

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 09:25:08 ; Search time 29.87 Seconds  
(without alignments)  
22.311 Million cell updates/sec

Title: BASK-853-CLAIM5  
Perfect score: 24  
Sequence: 1 gtxps 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 segs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	22	91.7	10 22 AAG94142	Human complementar
2	22	91.7	10 22 AAG97971	Human complementar
3	22	91.7	10 22 AAG97972	Human complementar
4	22	91.7	10 22 AAG97973	Human complementar
5	22	91.7	10 22 AAG97974	Human complementar
6	22	91.7	10 22 AAG97975	Human complementar
7	22	91.7	38 15 AAG58383	TSAR binding domain
8	22	91.7	50 22 AAU50435	Protonibacterium
9	22	91.7	52 21 AAB53513	Human colon cancer
10	22	91.7	53 22 AAU47468	Protonibacterium
11	22	91.7	58 22 AAU48691	Protonibacterium

12	22	91.7	59	22	AAO10898	Human polypeptide
13	22	91.7	61	22	AAU62078	Protonibacterium
14	22	91.7	62	22	AAU50474	Protonibacterium
15	22	91.7	76	22	AAU61979	Protonibacterium
16	22	91.7	78	20	AAU29072	T. gondii immunoge
17	22	91.7	78	22	AAU25543	T. gondii immunoge
18	22	91.7	80	22	AAU40914	Protonibacterium
19	22	91.7	82	22	AAU49223	Protonibacterium
20	22	91.7	86	22	AAG76165	Human colon cancer
21	22	91.7	90	22	AAU90260	Human immune/haema
22	22	91.7	91	22	ABE39057	Peptide #6563 enco
23	22	91.7	91	22	AAU95131	Human reproductive
24	22	91.7	91	22	AAU59709	Human brain expres
25	22	91.7	91	22	AAU72289	Human bone marrow
26	22	91.7	91	22	AAU32548	Peptide #6585 enco
27	22	91.7	106	22	AAU90809	Human immune/haema
28	22	91.7	109	21	AAU24569	Arabidopsis thalia
29	22	91.7	110	22	AAU24569	Human digestive sy
30	22	91.7	116	20	AAU12647	Human 5' EST secre
31	22	91.7	118	22	AAU65364	Protonibacterium
32	22	91.7	122	22	AAU50960	Protonibacterium
33	22	91.7	157	22	AAU92476	Human protein sequ
34	22	91.7	160	21	AAU32650	Eucalyptus grandis
35	22	91.7	160	21	AAU24568	Arabidopsis thalia
36	22	91.7	166	18	AAU32476	BBC6 protein for r
37	22	91.7	168	17	AAU95975	Nucleic acid recog
38	22	91.7	173	22	AAU38749	Novel subtilisin h
39	22	91.7	173	22	AAU38776	Novel subtilisin h
40	22	91.7	173	22	AAU38780	Novel subtilisin h
41	22	91.7	173	22	AAU38787	Novel subtilisin h
42	22	91.7	173	22	AAU38804	Novel subtilisin h
43	22	91.7	173	22	AAU38810	Novel subtilisin h
44	22	91.7	173	22	AAU38835	Novel subtilisin h
45	22	91.7	173	22	AAU38843	Novel subtilisin h
46	22	91.7	173	22	AAU38869	Novel subtilisin h
47	22	91.7	175	22	ABE68003	Drosophila melanog
48	22	91.7	182	22	ABE68002	Drosophila melanog
49	22	91.7	190	22	AAU40840	Human polypeptide
50	22	91.7	194	22	AAU40065	Protonibacterium

## ALIGNMENTS

RESULT 1  
AAG94142  
ID AAG94142 standard; Peptide; 10 AA.

XX AC AAG94142;

DT 18-SEP-2001 (first entry)

DE Human complementary peptide, SEQ ID NO: 336.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.

XX WO200142277-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.

XX 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI: 2001-408419/43.

PT A set of peptide ligands consisting of specific complementary peptides

PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -

PS Example 4; Page 88; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;

Query Match 91.7%; Score 22; DB 22; Length 10;  
 Best Local Similarity 66.7%; Pred. No. 47;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 5 gtssps 10

# RESULT 2

AAG97971  
 ID AAG97971 standard; Peptide; 10 AA.

AC AAG97971;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 4166.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.

XX WO200142277-A2.

PD 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.

XX 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -

XX Example 6; Page 630; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;

Query Match 91.7%; Score 22; DB 22; Length 10;  
 Best Local Similarity 66.7%; Pred. No. 47;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 5 gtssps 10

# RESULT 3

AAG97972  
 ID AAG97972 standard; Peptide; 10 AA.

XX AAG97972;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 4167.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.

XX WO200142277-A2.

XX 14-JUN-2001.

XX 13-DEC-2000; 2000WO-GB04776.

XX 13-DEC-1999; 99GB-0029464.

XX (PROT-) PROTEOM LTD.

XX Roberts GW, Heal JR;

XX WPI; 2001-408419/43.

XX A set of peptide ligands consisting of specific complementary peptides  
 PT to proteins encoded by genes of the human genome, useful in an assay  
 PT for screening and identifying of one or more novel peptides which are  
 PT drug candidates or pro-drugs -

XX Example 6; Page 630; 646pp; English.

XX The invention relates to a set of complementary peptide ligands  
 CC generated from the human genome. The complementary peptides  
 CC interact with their relevant target proteins encoded in the human  
 CC genome. They can be used as reagents in drug discovery and as lead  
 CC ligands to facilitate drug design and development. The present  
 CC sequence is a complementary peptide provided in the specification.

XX Sequence 10 AA;

Query Match 91.7%; Score 22; DB 22; Length 10;  
 Best Local Similarity 66.7%; Pred. No. 47;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 5 gtssps 10

# RESULT 4

AAG97973  
 ID AAG97973 standard; Peptide; 10 AA.

XX AAG97973;

XX 18-SEP-2001 (first entry)

XX Human complementary peptide, SEQ ID NO: 4168.

XX Human; complementary peptide; ligand; drug discovery; drug design.

XX Homo sapiens.



```
Db      4 gtsps 9
RESULT  7
AAR58383
ID      AAR58383 standard; Protein; 38 AA.
XX
XX
XX      AAR58383;
XX
XX      12-APR-1995 (first entry)
XX
XX      TSAR binding domain encoded by clone Zn1A7, binds zinc.
DE
XX      TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW      effector domain; concatenated heterofunctional protein; linker;
KW      direct; rapid; detection; screening; treatment; zinc binding.
XX
XX      Synthetic.
OS
XX
XX      WO9418318-A.
PN
XX
XX      18-AUG-1994.
PD
XX
XX      01-FEB-1994; 94WO-US00977.
PF
XX
XX      01-FEB-1993; 93US-0013416.
PR
XX      30-DEC-1993; 93US-0176500.
PR
XX      31-JAN-1994; 94US-0189331.
PR
XX
XX      (UYNC-) UNIV NORTH CAROLINA.
PA
XX
XX      Fowles DM, Kay BK;
PI
XX
XX      WPI: 1994-279739/34.
DR
XX      N-PSDB; AAQ70472.
DR
XX
XX      Identifying proteins or peptide(s) which bind a ligand - by
PT      screening a recombinant vector library expressing fusion proteins
PT      comprising a binding domain and an effector domain
XX
XX      Claim 38; Page 95; 255pp; English.
PS
XX
XX      AAR58376-94 show the amino acid sequences of the binding domain of TSAR
CC      (Totally Synthetic Affinity Reagents) peptides from the TSAR-9 library.
CC      These particular examples bind zinc. The non-variable amino acids at the
CC      NH2 and COOH terminals are not shown. TSAR peptides are generated using
CC      generic oligonucleotides (see AAQ70470-73 for examples). TSARs are
CC      concatenated heterofunctional proteins or peptides, comprising at least
CC      two functional regions - a binding domain with affinity for a ligand and
CC      a second effector peptide portion that is chemically or biologically
CC      active. They may further comprise a linker peptide between the 2 domains.
CC      The TSARs or compsns. comprising a TSAR binding domain can be used in
CC      vivo to deliver a chemically or biologically active moiety, eg. metal
CC      ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC      on the cell. They can also replace the function of macromolecules, eg.
CC      monoclonal or polyclonal antibodies and therefore circumvent the need
CC      for complex methods of hybridoma formation or in vivo antibody
CC      production. The TSARs are easily characterised and have designed
CC      activity allowing direct and rapid detection in a screening process.
XX
XX      Sequence 38 AA;
SQ
Query Match      91.7%; Score 22; DB 15; Length 38;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 gtxxps 6
      || ||
Db      20 gtsps 25
RESULT  8
AAU50435
ID      AAU50435 standard; Protein; 50 AA.
XX
XX      AAU50435;
AC
XX
XX      27-FEB-2002 (first entry)
DT
XX
XX      Propionibacterium acnes immunogenic protein #11331.
DE
XX
XX      SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW      uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW      inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW      dermatological; osteopathic; neuroprotectant.
XX
XX      Propionibacterium acnes.
OS
XX
XX      WO200181581-A2.
PN
XX
XX      01-NOV-2001.
PD
XX
XX      20-APR-2001; 2001WO-US12865.
PF
XX
XX      21-APR-2000; 2000US-199047P.
PR
XX      02-JUN-2000; 2000US-208841P.
PR
XX      07-JUL-2000; 2000US-216747P.
PR
XX
XX      (CORI-) CORIXA CORP.
PA
XX
XX      Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI      L'maisonneuve J, Zhang Y, Jen S, Carter D;
PI
XX
XX      WPI: 2001-616774/71.
DR
XX      N-PSDB; AAS59548.
DR
XX
XX      Propionibacterium acnes polypeptides and nucleic acids useful for
PT      vaccinating against and diagnosing infections, especially useful for
PT      treating acne vulgaris -
XX
XX      Example 1; SEQ ID No 11630; 1069pp; English.
PS
XX
XX      Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC      polypeptides. The proteins and their associated DNA sequences are used in
CC      the treatment, prevention and diagnosis of medical conditions caused by
CC      P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC      pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC      P. acnes is also involved in infections of bone, joints and the central
CC      nervous system, however it is particularly involved in the inflammatory
CC      lesions associated with acne vulgaris. A method for detecting the
CC      presence or absence of P. acnes in a patient comprises contacting a
CC      sample with a binding agent that binds to the proteins of the invention
CC      and determining the amount of bound protein in the sample. The
CC      polypeptides may be used as antigens in the production of antibodies
CC      specific for P. acnes proteins. These antibodies can be used to
CC      downregulate expression and activity of P. acnes polypeptides and
CC      therefore treat P. acnes infections. The antibodies may also be used as
CC      diagnostic agents for determining P. acnes presence, for example, by
CC      enzyme linked immunosorbent assay (ELISA).
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 50 AA;
SQ
Query Match      91.7%; Score 22; DB 22; Length 50;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 gtxxps 6
      || ||
Db      15 gtsps 20
```



ID AAU48691 standard; Protein; 58 AA.  
AC AAU48691;  
DT 27-FEB-2002 (first entry)  
DE  
DE Propionibacterium acnes immunogenic protein #9587.  
DE  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001WO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
PR 02-JUN-2000; 2000US-208841P.  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
DR WPI; 2001-616774/71.  
DR N-PSDB; AAS59543.  
XX  
PT Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
PS Example 1; SEQ ID No 9886; 1069pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 58 AA;  
  
Query Match 91.7%; Score 22; DB 22; Length 58;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 gtxxps 6  
DB 6 gtxxps 11  
  
RESULT 12

AAO10898  
ID AAO10898 standard; Protein; 59 AA.  
XX  
AC AAO10898;  
XX  
DT 06-NOV-2001 (first entry)  
DE  
DE Human polypeptide SEQ ID NO 24790.  
XX  
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
PN WO200164835-A2.  
XX  
PD 07-SEP-2001.  
XX  
PF 26-FEB-2001; 2001WO-US04927.  
XX  
PR 28-FEB-2000; 2000US-0515126.  
PR 18-MAY-2000; 2000US-0577409.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-514838/56.  
DR N-PSDB; AAI90829.  
XX  
PT Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
PS Claim 20; SEQ ID NO 24790; 1399pp + Sequence Listing; English.  
XX  
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 59 AA;  
  
Query Match 91.7%; Score 22; DB 22; Length 59;  
Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 gtxxps 6  
DB 49 gtxxps 54  
  
RESULT 13  
AAU62078  
ID AAU62078 standard; Protein; 61 AA.  
XX  
AC AAU62078;  
XX  
DT 27-FEB-2002 (first entry)  
XX

DE Propionibacterium acnes immunogenic protein #22974.  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59623.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 23273; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 61 AA;  
 XX  
 Query Match 91.7%; Score 22; DB 22; Length 61;  
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gtxxps 6  
 || ||  
 Db 42 gtsps 47  
 RESULT 14  
 AAU50474  
 ID AAU50474 standard; Protein; 62 AA.  
 XX  
 AC AAU50474;  
 XX  
 DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #11370.  
 DE  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59548.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 11669; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 62 AA;  
 XX  
 Query Match 91.7%; Score 22; DB 22; Length 62;  
 Best Local Similarity 66.7%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gtxxps 6  
 || ||  
 Db 43 gtsps 48  
 RESULT 15  
 AAU61979  
 ID AAU61979 standard; Protein; 76 AA.  
 XX  
 AC AAU61979;  
 XX

DT 27-FEB-2002 (first entry)  
XX Propionibacterium acnes immunogenic protein #22875.  
DE  
XX  
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acnes.  
XX  
XX WO200181581-A2.  
PN  
XX  
PD 01-NOV-2001.  
XX  
XX 20-APR-2001; 2001WO-US12865.  
PF  
XX 21-APR-2000; 2000US-199047P.  
PR  
PR 02-JUN-2000; 2000US-208841P.  
PR  
PR 07-JUL-2000; 2000US-216747P.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
PI  
XX WPI; 2001-616774/71.  
DR  
DR N-PSDB; AAS59621.  
XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
PT  
XX Example 1; SEQ ID No 23174; 1069pp; English.  
PS  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 76 AA;  
  
Query Match 91.7%; Score 22; DB 22; Length 76;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 2 gttssps 7  
  
RESULT 16  
AAU29072  
ID AAU29072 standard; Protein; 78 AA.  
XX  
AC AAU29072;  
  
Query Match 91.7%; Score 22; DB 22; Length 76;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 2 gttssps 7  
  
RESULT 16  
AAU29072  
ID AAU29072 standard; Protein; 78 AA.  
XX  
AC AAU29072;

XX 24-SEP-1999 (first entry)  
DT  
XX  
DE T. gondii immunogenic protein.  
XX  
KW Immunogenic protein; Toxoplasma gondii protein; oocyst shedding; cat;  
KW T. gondii infection; enteric apicomplexa oocyst; Cryptosporidium oocyst;  
KW Toxoplasma oocyst.  
XX  
OS Toxoplasma gondii.  
XX  
XX WO9932633-A1.  
PN  
XX  
PD 01-JUL-1999.  
XX  
XX 18-DEC-1998; 98WO-US27137.  
PF  
XX 19-DEC-1997; 97US-0994825.  
PR  
XX (HESK-) HESKA CORP.  
PA  
XX Lutz SB, Milhausen MJ, Ng RK;  
PI  
XX WPI; 1999-418930/35.  
DR  
DR N-PSDB; AAX91398.  
XX  
XX New isolated Toxoplasma gondii nucleic acids used, e.g. to treat  
PT infection caused by this microorganism  
PT  
XX Claim 29; Page 319-320; 381pp; English.  
PS  
XX The invention provides isolated Toxoplasma gondii nucleic acids that  
CC encode immunogenic polypeptides. The T. gondii nucleic acid molecules,  
CC immunogenic proteins and antibodies to the proteins can be used to  
CC inhibit T. gondii oocyst shedding in a cat due to infection with  
CC T. gondii. They can be used for preventing T. gondii infection and for  
CC preventing the spread of T. gondii infection. They can also be used for  
CC detecting T. gondii infection. The detection method can be used to detect  
CC parasite cysts or oocysts in feces, e.g. from enteric apicomplexa oocysts  
CC such as Cryptosporidium oocysts and Toxoplasma oocysts.  
XX  
SQ Sequence 78 AA;  
  
Query Match 91.7%; Score 22; DB 20; Length 78;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 50 gtspps 55  
  
RESULT 17  
AAU25543  
ID AAU25543 standard; Protein; 78 AA.  
XX  
XX AAU25543;  
AC  
XX  
XX 17-DEC-2001 (first entry)  
DT  
XX  
DE T. gondii immunogenic protein PM2A18-a.  
XX  
KW Immunogenic protein; oocyst; faeces; enteric apicomplexa oocyst;  
KW Cryptosporidium oocyst; Toxoplasma oocyst; Giardia cyst; vaccine;  
KW oocyte shedding.  
XX  
OS Toxoplasma gondii.  
XX  
XX US2001014447-A1.  
PN  
XX  
PD 16-AUG-2001.  
XX

PF 18-DEC-1998; 98US-0216393.  
 XX  
 PR 19-DEC-1997; 97US-0994825.  
 XX  
 PA (MTLH/) MILHAUSEN M J.  
 XX  
 PI Milhausen MJ;  
 XX  
 DR WPI: 2001-529100/58.  
 DR N-PSDB; AAS42721.  
 XX  
 XX Detecting parasite oocysts or cysts in faeces, comprises eluting DNA  
 PT from sample into aqueous solution by heating, amplifying DNA with  
 CC primers specific for oocysts or cysts being detected, and detecting  
 PT amplification product.  
 XX  
 PS Disclosure: Page 150; 188pp; English.  
 XX  
 CC The invention relates to detection of parasite oocysts or cysts in  
 CC a faeces sample comprising contacting the sample with a solid support,  
 CC drying and then washing the sample with an aqueous wash solution, adding  
 CC an aqueous elution solution and eluting DNA from the sample by heating  
 CC and amplifying by PCR oocyst/cyst-specific DNA and detecting the  
 CC amplification products. The method is useful for detecting parasite  
 CC oocysts e.g., enteric apicomplexa oocysts such as Cryptosporidium oocysts  
 CC or Toxoplasma oocysts, or for detecting parasite cysts e.g. Giardia  
 CC cysts. The method is also useful for developing vaccines to prevent  
 CC oocyte shedding in cats. The present sequence represents an  
 CC immunogenic protein from Toxoplasma gondii.  
 XX  
 SQ Sequence 78 AA;

Query Match 91.7%; Score 22; DB 22; Length 78;  
 Best Local Similarity 66.7%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 50 gtspps 55

RESULT 18  
 AAU40914  
 ID AAU40914 standard; Protein; 80 AA.  
 XX  
 AC AAU40914;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #1810.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI: 2001-616774/71.  
 DR N-PSDB; AAS59513.  
 XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -

XX Example 1; SEQ ID No 2109; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA)  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 80 AA;

Query Match 91.7%; Score 22; DB 22; Length 80;  
 Best Local Similarity 66.7%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 22 gtspps 27

RESULT 19  
 AAU49223  
 ID AAU49223 standard; Protein; 82 AA.  
 XX  
 AC AAU49223;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #10119.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
DR N-PSDB; AAS59545.  
XX  
XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
XX Example 1; SEQ ID NO 10418; 1069pp; English.  
XX  
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 82 AA;

Query Match 91.7%; Score 22; DB 22; Length 82;  
Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtxxps 6  
Db 33 gttaps 38  
|||  
|||

RESULT 20  
AAG76165  
ID AAG76165 standard; Protein; 86 AA.  
XX  
XX AAG76165;  
AC  
XX 03-SEP-2001 (first entry)  
DT  
XX Human colon cancer antigen protein SEQ ID NO:6929.  
DE  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; chromosome 12.  
KW  
XX Homo sapiens.  
OS  
XX WO200122920-A2.  
PN  
XX 05-APR-2001.  
PD  
XX 28-SEP-2000; 2000WO-US26524.  
PF  
XX 29-SEP-1999; 99US-0157137.  
PR  
XX 03-NOV-1999; 99US-0163280.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
PI WPI; 2001-235357/24.  
XX

DR N-PSDB; AAH35570.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
XX Claim 11; Page 8369; 9803pp; English.  
PS  
XX AAH32943 to AAH37195 and AAG73514 to AAG77798 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC and AAH77789 represent sequences used in the exemplification of the  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were for  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 86 AA;

Query Match 91.7%; Score 22; DB 22; Length 86;  
Best Local Similarity 66.7%; Pred. No. 3.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtxxps 6  
Db 37 gttaps 42  
|||  
|||

RESULT 21  
AAM90260  
ID AAM90260 standard; Protein; 90 AA.  
XX  
XX AAM90260;  
AC  
XX 07-NOV-2001 (first entry)  
DT  
XX Human immune/haematopoietic antigen SEQ ID NO:17853.  
DE  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
KW  
XX Homo sapiens.  
OS  
XX WO200157182-A2.  
PN  
XX 09-AUG-2001.  
PD  
XX 17-JAN-2001; 2001WO-US01354.  
PF  
XX 31-JAN-2000; 2000US-0179065.  
PR  
XX 04-FEB-2000; 2000US-0180628.  
PR  
XX 24-FEB-2000; 2000US-0184664.  
PR  
XX 02-MAR-2000; 2000US-0186350.  
PR  
XX 16-MAR-2000; 2000US-0189874.  
PR  
XX 17-MAR-2000; 2000US-0190076.  
PR  
XX 18-APR-2000; 2000US-0198123.  
PR  
XX 19-MAY-2000; 2000US-0205515.  
PR  
XX 28-JUN-2000; 2000US-0209467.  
PR  
XX 30-JUN-2000; 2000US-0214886.  
PR  
XX 07-JUL-2000; 2000US-0215135.  
PR  
XX 07-JUL-2000; 2000US-0216647.  
PR  
XX 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232357.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0251989.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI: 2001-483426/52.  
N-PSDB: AAK63041.  
Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -  
Claim 11; SEQ ID NO 17853; 3071pp + Sequence Listing; English.  
AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients own production of (I). Additionally, (I)

CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (i) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.

XX Sequence 90 AA;

Query Match 91.7%; Score 22; DB 22; Length 90;  
Best Local Similarity 66.7%; Pred. NO. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtxxps 6  
|| ||  
Db 1 gtasps 6

RESULT 22

ABB39057  
ID ABB39057 standard; Peptide; 91 AA.

XX AC ABB39057;

XX DT 04-FEB-2002 (first entry)

XX DE Peptide #6563 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157277-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00669.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207450.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XW WPI; 2001-483447/52.

XX DR Human genome-derived single exon nucleic acid probes useful for

XX PT analyzing gene expression in human fetal liver -

XX PS Claim 27; SEQ ID NO 31692; 639pp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for

XX CC measuring human gene expression in a sample derived from human foetal

XX CC liver. The single exon nucleic acid probes may be used for predicting,

XX CC measuring and displaying gene expression in samples derived from human

XX CC fetal liver. The present sequence is a peptide encoded by a single exon

XX CC nucleic acid probe of the invention.

XX CC Note: The sequence data for this patent did not form part of the

XX CC printed specification, but was obtained in electronic format directly

XX CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 91 AA;

Query Match 91.7%; Score 22; DB 22; Length 91;  
Best Local Similarity 66.7%; Pred. NO. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtxxps 6  
|| ||  
Db 56 gtssps 61

RESULT 23

AAM95131  
ID AAM95131 standard; Protein; 91 AA.

XX AC AAM95131;

XX DT 21-NOV-2001 (first entry)

XX DE Human reproductive system related antigen SEQ ID NO: 3789.

XX KW Human; reproductive system related antigen; reproductive system disorder;

XX KW cancer; gene therapy.

XX OS Homo sapiens.

XX PN WO200155320-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01339.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226686.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229345.

XX PR 05-SEP-2000; 2000US-0229509.

XX PR 06-SEP-2000; 2000US-0229513.

XX PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234423.  
PR 21-SEP-2000; 2000US-0234474.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0241821.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-024617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
DA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
DR N-PSDB; AAL01101.  
XX  
PT Isolated nucleic acid molecule encoding a reproductive system antigen -  
PT is used in preventing, treating or ameliorating a medical condition -  
XX  
PS Claim 11; SEQ ID NO 3789; 1297pp + Sequence Listing; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a protein of the invention.  
XX  
SQ Sequence 91 AA;  
  
Query Match 91.7%; Score 22; DB 22; Length 91;  
Best Local Similarity 66.7%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 gtxxps 6  
|| ||  
Db 16 gstats 21  
  
RESULT 24  
ARM59709  
ID AAM59709 standard; Protein; 91 AA.  
XX  
AC AAM59709;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31814.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WC200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483446/52.  
DR  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
PT  
XX  
XX Example 4; SEQ ID NO: 31814; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
XX Sequence 91 AA;  
SQ

Query Match 91.7%; Score 22; DB 22; Length 91;  
Best Local Similarity 66.7%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 56 gtsps 61

RESULT 25  
AAM72289  
ID AAM72289 standard; Protein; 91 AA.  
XX  
XX AC AAM72289;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32595.  
XX  
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157276-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00668.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX  
XX PR 26-MAY-2000; 2000US-0207456.  
XX  
XX PR 30-JUN-2000; 2000US-0608408.  
XX  
XX PR 03-AUG-2000; 2000US-0632366.  
XX  
XX PR 21-SEP-2000; 2000US-0234687.  
XX  
XX PR 27-SEP-2000; 2000US-0236359.  
XX  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX  
XX

PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
XX Example 4; SEQ ID NO: 32595; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention.  
XX  
XX Sequence 91 AA;  
SQ

Query Match 91.7%; Score 22; DB 22; Length 91;  
Best Local Similarity 66.7%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 56 gtsps 61

RESULT 26  
AAM32548  
ID AAM32548 standard; Protein; 91 AA.  
XX  
XX AC AAM32548;  
XX  
XX DT 17-OCT-2001 (first entry)  
XX  
XX DE Peptide #6585 encoded by probe for measuring placental gene expression.  
XX  
XX KW Probe; microarray; human; placenta; antenatal diagnosis;  
KW genetic disorder.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157272-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 30-JAN-2001; 2001WO-US00663.  
XX  
XX PR 04-FEB-2000; 2000US-0180312.  
XX  
XX PR 26-MAY-2000; 2000US-0207456.  
XX  
XX PR 30-JUN-2000; 2000US-0608408.  
XX  
XX PR 03-AUG-2000; 2000US-0632366.  
XX  
XX PR 21-SEP-2000; 2000US-0234687.  
XX  
XX PR 27-SEP-2000; 2000US-0236359.  
XX  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX  
XX DE Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human placenta -  
XX  
XX PS Claim 27; SEQ ID No 32817; 654pp; English.  
XX  
XX CC The present invention relates to single exon nucleic acid probes (SENP;  
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one  
XX such probe. The probes are useful for producing a microarray for  
XX predicting, measuring and displaying gene expression in samples derived  
XX from human placenta. The probes are useful for antenatal diagnosis of  
XX human genetic disorders.  
XX  
XX Sequence 91 AA;  
SQ

```
Query Match          91.7%; Score 22; DB 22; Length 91;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 56 gtsps 61

RESULT 27
AAM90809
ID AAM90809 standard; Protein; 106 AA.
AC
XX
XX
DT
XX
DE Human immune/haematopoietic antigen SEQ ID NO:18402.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD
XX
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
```

PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX N-PSDB; AAK63590.  
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX Claim 11; SEQ ID NO 18402; 3071pp + Sequence Listing; English.  
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX Sequence 106 AA;  
SQ

Query Match 91.7%; Score 22; DB 22; Length 106;  
Best Local Similarity 66.7%; Pred. No. 4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 gtxxps 6  
|| ||  
Db 37 gtsaps 42

RESULT 28  
AAG24569  
ID AAG24569 standard; Protein; 109 AA.  
XX  
AC AAG24569;  
XX  
DT 17-OCT-2000 (first entry)  
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 28293.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS EPI033405-A2.  
PN 06-SEP-2000.  
PD 25-FEB-2000; 2000EP-0301439.  
PF 25-FEB-1999; 99US-0121825.  
XX 09-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.

```
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
```

```
PR 30-AUG-1999; 99US-01511303.
PR 31-AUG-1999; 99US-01511438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 91.7%; Score 22; DB 21; Length 109;  
Best Local Similarity 66.7%; Pred. No. 4.1e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|||  
Db 71 gtttpps 76

RESULT 29  
AAM92684  
ID AAM92684 standard; Protein; 110 AA.

XX AC AAM92684;

XX DT 06-NOV-2001 (first entry)

XX Human digestive system antigen SEQ ID NO: 2033.

KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
XX digestive system disorder; Meckel's diverticulum.  
OS Homo sapiens.

XX WO20015314-A2.  
PN 02-AUG-2001.  
PD 17-JAN-2001; 2001WO-US01324.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250161.  
PR 05-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX

PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-502630/55.  
 DR N-PSDB; AAK88457.  
 XX  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases -  
 XX  
 XX Claim 11; SEQ ID NO 2033; 986pp; English.  
 XX  
 XX The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a digestive system antigen of  
 CC the invention.  
 XX  
 SQ Sequence 110 AA;  
 XX  
 Query Match 91.7%; Score 22; DB 22; Length 110;  
 Best Local Similarity 66.7%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gtxxps 6  
 || ||  
 Db 13 gtsps 18  
 || ||  
 RESULT 30  
 AAY12647  
 ID AAY12647 standard; Protein; 116 AA.  
 XX  
 AC AAY12647;  
 XX  
 XX 22-JUN-1999 (first entry)  
 DT  
 XX  
 DE Human 5' EST secreted protein SEQ ID NO: 312 from WO 9906553.  
 XX  
 XX Human; secreted protein; EST; expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; antiinflammatory; tumour inhibition; antitumour.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO9906553-A2.  
 PN  
 XX 11-FEB-1999.  
 PD  
 XX 31-JUL-1998; 98WO-IB01237.  
 PF  
 XX 01-AUG-1997; 97US-0905051.  
 XX  
 PR (GEST ) GENSET.  
 PA  
 XX  
 XX Duclert A, Dumas Milne Edwards J, Lacroix B;  
 PI  
 XX WPI; 1999-153783/13.  
 DR N-PSDB; AAX41505.  
 DR  
 XX New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries derived from umbilical cord, lymph ganglia,  
 PT lymphocytes and placental tissue  
 PT  
 XX  
 XX Claim 34; Page 397-398; 411pp; English.  
 PS  
 XX AAX41379 to AAX41526 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins, and encode the proteins given in AAY12521 to  
 CC

CC AAY12668, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/ chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ ligand activity, antiinflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 XX  
 SQ Sequence 116 AA;  
 XX  
 Query Match 91.7%; Score 22; DB 20; Length 116;  
 Best Local Similarity 66.7%; Pred. No. 4.4e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gtxxps 6  
 || ||  
 Db 53 gtsaps 58  
 || ||  
 RESULT 31  
 AAU65364  
 ID AAU65364 standard; Protein; 118 AA.  
 XX  
 AC AAU65364;  
 XX  
 DT 27-FEB-2002 (first entry)  
 DT  
 XX  
 DE Propionibacterium acnes immunogenic protein #26260.  
 XX  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 XX WO200181581-A2.  
 PN  
 XX 01-NOV-2001.  
 PD  
 XX 20-APR-2001; 2001WO-US12865.  
 XX  
 XX 21-APR-2000; 2000US-199047P.  
 PR  
 XX 02-JUN-2000; 2000US-208841P.  
 PR  
 XX 07-JUL-2000; 2000US-216747P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 PA  
 XX  
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 PI  
 XX WPI; 2001-616774/71.  
 DR N-PSDB; AAS59666.  
 DR  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 PT  
 XX Example 1; SEQ ID No 26559; 1069pp; English.  
 PS  
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 118 AA;

Query Match 91.7%; Score 22; DB 22; Length 118;  
 Best Local Similarity 66.7%; Pred. No. 4.4e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 111 gtatps 116

# RESULT 32

AAU50960  
 ID AAU50960 standard; Protein; 122 AA.

XX AAU50960;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #11856.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001WO-US12865.

PR 21-APR-2000; 2000US-195047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

PA (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX N-PSDB; AAS59549.

DR WPI; 2001-616774/71.

XX Example 1; SEQ ID No 12155; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by

CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 122 AA;

Query Match 91.7%; Score 22; DB 22; Length 122;  
 Best Local Similarity 66.7%; Pred. No. 4.6e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 18 gtttps 23

# RESULT 33

AAAB92476  
 ID AAAB92476 standard; Protein; 157 AA.

XX AAAB92476;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:10552.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -

XX Claim 8; SEQ ID 10552; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesising polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX SQ Sequence 157 AA;

Query Match 91.7%; Score 22; DB 22; Length 157;

Best Local Similarity 66.7%; Pred. No. 5.7e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6

Db 76 gtaaps 81

RESULT 34

AAB32650

ID AAB32650 standard; Protein; 160 AA.

XX AC AAB32650;

DT 25-JAN-2001 (first entry)

XX DE Eucalyptus grandis transcription factor protein sequence #108.

XX KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
 KW homeodomain zipper; LIM domain; AP2; EREs; zinc finger domain;  
 KW type 2 Cys2His2; CCAAT box element; MYB.

XX OS Eucalyptus grandis.

XX PN WO200053724-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06112.

XX PR 11-MAR-1999; 99US-0266513.

XX PR 18-AUG-1999; 99US-0149485.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Wood M, McGrath A, Shenk MA, Glenn M;

XX DR WPI; 2000-579369/54.

XX DR

XX PT New isolated polynucleotide encoding a plant transcription factor for  
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
 PT having modified gene expression or modified activity of a polypeptide

XX PS Claim 8; Page 243; 747pp; English.

XX

CC The present invention relates to novel plant transcription factors from  
 CC Eucalyptus grandis or Pinus radiata. The present sequence is one such  
 CC transcription factor. The transcription factor may be used to produce a  
 CC plant having modified gene expression such as a woody plant e.g. a  
 CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
 CC to modify the activity of a polypeptide in a plant. The transcription  
 CC factors of the present invention are members from the following families  
 CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
 CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
 CC zipper, LIM domain, AP2 and EREs, zinc finger domains of type 2  
 CC Cys2His2, CCAAT box elements and MYB.

XX SQ Sequence 160 AA;

Query Match 91.7%; Score 22; DB 21; Length 160;

Best Local Similarity 66.7%; Pred. No. 5.8e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6

Db 112 gtaaps 117

RESULT 35

AAG24568

ID AAG24568 standard; Protein; 160 AA.

XX AC AAG24568;

DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 28292.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 04-MAY-1999; 99US-0132407.

XX PR 05-MAY-1999; 99US-0132484.

XX PR 06-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 07-MAY-1999; 99US-0132487.

XX PR 11-MAY-1999; 99US-0132863.

XX PR 14-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151086.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161921.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 91.7%; Score 22; DB 21; Length 160;  
Best Local Similarity 66.7%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 122 gtttpps 127

RESULT 36  
AAW32476  
ID AAW32476 standard; Protein; 166 AA.

XX AC AAW32476;  
XX DT 15-JAN-1998 (first entry)  
XX DE BBC6 protein for regulating cell death.

XX KW BBC6 gene; cell death; cell cycle; Bcl2; human.  
XX OS Homo sapiens.

XX PN US5663316-A.  
XX PD 02-SEP-1997.  
XX PF 18-JUN-1996; 96US-0665617.

XX PR 18-JUN-1996; 96US-0665617.  
XX PA (CLON-) CLONTECH LAB INC.  
XX PI Xudong Y;

XX DR WPI; 1997-447980/41.  
XX DR N-PSDB; AAT91561.

XX PT Isolated BBC6 gene - encodes a protein that regulates cell death  
XX PT through interaction with Bcl-2

PS Claim 1; Column 11-12; 7pp; English.

XX The present sequence represents a protein of 166 amino acids. The  
CC sequence is disclosed as being a protein called BBC6 which regulates  
CC cell death through interaction with Bcl-2. The DNA may be used for the  
CC production of the recombinant protein, which can be used in unspecified  
CC therapeutic or diagnostic procedures, as a molecular weight marker, and  
CC to raise antibodies that can be used in unspecified diagnostic or  
CC therapeutic applications and to reduce or eliminate the biological  
CC activity of the BBC6 protein in vivo.

XX Sequence 166 AA;

Query Match 91.7%; Score 22; DB 18; Length 166;  
Best Local Similarity 66.7%; Pred. No. 6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 160 gtaaps 165

RESULT 37

AAW95975  
ID AAR95975 standard; peptide; 168 AA.

XX AC AAR95975;  
XX DT 19-FEB-1997 (first entry)  
XX DE Nucleic acid recognition unit #11.

XX KW Target binding assembly; nucleic acid recognition unit; NF-kappa-B; SPL;

XX KW TATA; human papillomavirus; HPV; HPV E2; human immunodeficiency virus;  
XX KW HIV LTR; Tat binding unit; probe; assembly sequence; asymmetry sequence;  
XX KW nuclear localisation signal sequence; human cell; HPV LTR; HIV; bacteria;  
XX virus.

XX OS Synthetic.

XX PN WO9617956-A2.

XX PD 13-JUN-1996.

XX PF 07-DEC-1995; 95WO-US15944.

XX PR 09-DEC-1994; 94US-0353476.

XX PA (GENE-) GENE POOL INC.

XX PI Weininger AM, Weininger S;

XX DR WPI; 1996-287199/29.

XX PT Probe nucleic acids, target binding assemblies, etc - for detection  
XX PT and localisation of specific nucleic acid sequences, esp. HIV and  
XX PT HPV

XX Claim 14; Page 97-98; 172pp; English.

XX AAR95965-R95993 represent the nucleic acid recognition units (NAR) of  
CC target binding assemblies (TBA) of the invention. These NARs are  
CC selected from NF-kappa-B, SPL, TATA, human papillomavirus (HPV) E2, HPV  
CC LTR, human immunodeficiency virus (HIV) LTR and Tat binding units. This  
CC sequence represents a SPL sequence NAR. The TBA is recognised by the  
CC target binding region (TBR) of a probe of the invention. The probe of  
CC the invention contains a TBR, a booster binding region (BBR), and an  
CC optional support or attachment (OSA). The TBA contains at least one  
CC NAR, and optionally a linker sequence, an assembly sequence, an  
CC asymmetry sequence, a nuclear localisation signal sequence, and an OSA.  
CC The assembly sequence and asymmetry sequences are responsible for the  
CC folding and association of the NARs. The linker sequence is an  
CC oligopeptide, which does not interfere with NAR function, but provides  
CC stability and control over the spacing of the NAR from the rest of the  
CC TBA. The OSA is an attached support or indicator, or other means of  
CC localisation of the probe. The probe can be used in a method for  
CC detecting or localising a specific target nucleic acid sequence (TNA).  
CC The method is highly sensitive, and has a high degree of specificity.  
CC The method can be used for detecting specific nucleic acid sequences,  
CC including those found in human cells, in HIV, HPV, and other nucleic acid  
CC containing systems, including bacteria and viruses.

XX Sequence 168 AA;

Query Match 91.7%; Score 22; DB 17; Length 168;  
Best Local Similarity 66.7%; Pred. No. 6.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 119 gtaaps 124

RESULT 38  
AAU38749

ID AUU38749 standard; Protein; 173 AA.  
AC AUU38749;  
XX  
XX  
DT 19-DEC-2001 (first entry)  
XX  
XX  
DE Novel subtilisin homologue #6.  
XX  
XX Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
KW contact lens cleansing solution; dry cleaning.  
XX  
XX Bacillus sp.  
XX  
XX WO200175087-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 02-APR-2001; 2001WO-US10781.  
XX  
XX 03-APR-2000; 2000US-194143P.  
XX  
XX (MAXY-) MAXYGEN INC.  
XX  
XX Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
PI Minshull J;  
XX  
XX WPI; 2001-616689/71.  
XX  
XX N-PSDB; AAS56604.  
XX  
XX Novel subtilisin homologue polypeptides having improved endo-protease  
PT activity relative to mature Savinase subtilisin polypeptide, useful as  
PT component of cleaning solutions e.g. laundry detergents, dry cleaning  
PT  
XX  
XX  
PS Claim 1; Page 103; 138pp; English.  
XX  
XX The invention relates to an isolated polypeptide with improved endo-  
CC protease activity relative to subtilisin homologue polypeptide, Savinase  
CC (RTM). The invention also relates to an integrated system comprising a  
CC computer or computer readable medium comprising a database comprising one  
CC or more sequence records. Each record comprises one or more character  
CC string corresponding to a nucleic acid or protein sequence of AAS56599-  
CC AAS56728 or AUU38744-AUU38873 respectively. The integrated system  
CC comprises a user input interface allowing a user to selectively  
CC one or more sequence record. The integrated system is useful for  
CC presenting information pertaining to one of several sequence records  
CC stored in a database. The method involves determining a list of one or  
CC more character strings corresponding to the sequence or its subsequence,  
CC and displaying the selected character strings or aligning the selected  
CC character string with an additional character string. The method  
CC further involves displaying an alignment of the selected character  
CC string with the additional character string, and displaying the list.  
CC The polynucleotides are useful as probes, primers, sense and antisense  
CC nucleotides. The subtilisin homologue polypeptides are useful for  
CC producing antibodies which have diagnostic uses related to the activity,  
CC distribution and expression of subtilisin homologues. The subtilisin  
CC homologue polypeptides are used in compositions that serve as cleaning  
CC solutions in a wide variety of applications including laundry detergents,  
CC contact lens cleansing solutions, and dry cleaning. AUU38744-AUU38874  
XX represent the amino acid sequences of novel subtilisin homologues of the  
XX invention.  
XX  
SQ Sequence 173 AA;

Query Match 91.7%; Score 22; DB 22; Length 173;  
Best Local Similarity 66.7%; Pred. No. 6.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

Db 71 gtsaps 76

RESULT 39  
AAU38776  
ID AUU38776 standard; Protein; 173 AA.  
XX  
XX  
AC AUU38776;  
XX  
XX  
DT 19-DEC-2001 (first entry)  
XX  
XX  
DE Novel subtilisin homologue #33.  
XX  
XX Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
KW contact lens cleansing solution; dry cleaning.  
XX  
XX Bacillus sp.  
XX  
XX WO200175087-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 02-APR-2001; 2001WO-US10781.  
XX  
XX 03-APR-2000; 2000US-194143P.  
XX  
XX (MAXY-) MAXYGEN INC.  
XX  
XX Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
PI Minshull J;  
XX  
XX WPI; 2001-616689/71.  
XX  
XX N-PSDB; AAS56631.  
XX  
XX Novel subtilisin homologue polypeptides having improved endo-protease  
PT activity relative to mature Savinase subtilisin polypeptide, useful as  
PT component of cleaning solutions e.g. laundry detergents, dry cleaning  
PT  
XX  
XX  
PS Claim 1; Page 106; 138pp; English.  
XX  
XX The invention relates to an isolated polypeptide with improved endo-  
CC protease activity relative to subtilisin homologue polypeptide, Savinase  
CC (RTM). The invention also relates to an integrated system comprising a  
CC computer or computer readable medium comprising a database comprising one  
CC or more sequence records. Each record comprises one or more character  
CC string corresponding to a nucleic acid or protein sequence of AAS56599-  
CC AAS56728 or AUU38744-AUU38873 respectively. The integrated system  
CC comprises a user input interface allowing a user to selectively  
CC one or more sequence record. The integrated system is useful for  
CC presenting information pertaining to one of several sequence records  
CC stored in a database. The method involves determining a list of one or  
CC more character strings corresponding to the sequence or its subsequence,  
CC and displaying the selected character strings or aligning the selected  
CC character string with an additional character string. The method  
CC further involves displaying an alignment of the selected character  
CC string with the additional character string, and displaying the list.  
CC The polynucleotides are useful as probes, primers, sense and antisense  
CC nucleotides. The subtilisin homologue polypeptides are useful for  
CC producing antibodies which have diagnostic uses related to the activity,  
CC distribution and expression of subtilisin homologues. The subtilisin  
CC homologue polypeptides are used in compositions that serve as cleaning  
CC solutions in a wide variety of applications including laundry detergents,  
CC contact lens cleansing solutions, and dry cleaning. AUU38744-AUU38874  
XX represent the amino acid sequences of novel subtilisin homologues of the  
XX invention.  
XX  
SQ Sequence 173 AA;

Query Match 91.7%; Score 22; DB 22; Length 173;  
Best Local Similarity 66.7%; Pred. No. 6.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 71 gtsaps 76

RESULT 40  
AAU38780  
ID AAU38780 standard; Protein; 173 AA.

XX AC AAU38780;  
XX DT 19-DEC-2001 (first entry)

XX DE Novel subtilisin homologue #37.

XX KW Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
KW contact lens cleansing solution; dry cleaning.

XX OS Bacillus sp.

XX PN WO200175087-A2.

XX PD 11-OCT-2001.

XX PF 02-APR-2001; 2001WO-US10781.

XX PR 03-APR-2000; 2000US-194143P.

XX PA (MAXY-) MAXYGEN INC.

XX PI Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
PI Minshull J;

XX DR WPI: 2001-616689/71.  
DR N-PSDB; AAS56635.

XX Novel subtilisin homologue polypeptides having improved endo-protease  
PT activity relative to mature Savinase subtilisin polypeptide, useful as  
PT component of cleaning solutions e.g. laundry detergents, dry cleaning  
PT -

PS Claim 1; Page 106; 138pp; English.

XX The invention relates to an isolated polypeptide with improved endo-  
CC protease activity relative to subtilisin homologue polypeptide, Savinase  
CC (RTM). The invention also relates to an integrated system comprising a  
CC computer or computer readable medium comprising a database comprising one  
CC or more sequence records. Each record comprises one or more character  
CC string corresponding to a nucleic acid or protein sequence of AAS56599-  
CC AAS56728 or AAU38744-AAU38873 respectively. The integrated system  
CC comprises a user input interface allowing a user to selectively  
CC one or more sequence record. The integrated system is useful for  
CC presenting information pertaining to one of several sequence records  
CC stored in a database. The method involves determining a list of one or  
CC more character strings corresponding to the sequence or its subsequence,  
CC determining which character strings of the list are selected by a user,  
CC and displaying the selected character strings or aligning the selected  
CC character string with an additional character string. The method  
CC further involves displaying an alignment of the selected character  
CC string with the additional character string, and displaying the list.  
CC The polynucleotides are useful as probes, primers, sense and antisense  
CC nucleotides. The subtilisin homologue polypeptides are useful for  
CC producing antibodies which have diagnostic uses related to the activity,  
CC distribution and expression of subtilisin homologues. The subtilisin  
CC homologue polypeptides are used in compositions that serve as cleaning  
CC solutions in a wide variety of applications including laundry detergents,  
CC contact lens cleansing solutions, and dry cleaning. AAU38744-AAU38874  
CC represent the amino acid sequences of novel subtilisin homologues of the  
CC invention.

XX Sequence 173 AA;

SQ

Query Match 91.7%; Score 22; DB 22; Length 173;  
Best Local Similarity 66.7%; Pred. No. 6.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 71 gtsaps 76

RESULT 41  
AAU38787  
ID AAU38787 standard; Protein; 173 AA.

XX AC AAU38787;

XX DT 19-DEC-2001 (first entry)

XX DE Novel subtilisin homologue #44.

XX KW Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
KW contact lens cleansing solution; dry cleaning.

XX OS Bacillus sp.

XX PN WO200175087-A2.

XX PD 11-OCT-2001.

XX PF 02-APR-2001; 2001WO-US10781.

XX PR 03-APR-2000; 2000US-194143P.

XX PA (MAXY-) MAXYGEN INC.

XX PI Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
PI Minshull J;

XX DR WPI: 2001-616689/71.  
DR N-PSDB; AAS56642.

XX Novel subtilisin homologue polypeptides having improved endo-protease  
PT activity relative to mature Savinase subtilisin polypeptide, useful as  
PT component of cleaning solutions e.g. laundry detergents, dry cleaning  
PT -

PS Claim 1; Page 107; 138pp; English.

XX The invention relates to an isolated polypeptide with improved endo-  
CC protease activity relative to subtilisin homologue polypeptide, Savinase  
CC (RTM). The invention also relates to an integrated system comprising a  
CC computer or computer readable medium comprising a database comprising one  
CC or more sequence records. Each record comprises one or more character  
CC string corresponding to a nucleic acid or protein sequence of AAS56599-  
CC AAS56728 or AAU38744-AAU38873 respectively. The integrated system  
CC comprises a user input interface allowing a user to selectively  
CC one or more sequence record. The integrated system is useful for  
CC presenting information pertaining to one of several sequence records  
CC stored in a database. The method involves determining a list of one or  
CC more character strings corresponding to the sequence or its subsequence,  
CC determining which character strings of the list are selected by a user,  
CC and displaying the selected character strings or aligning the selected  
CC character string with an additional character string. The method  
CC further involves displaying an alignment of the selected character  
CC string with the additional character string, and displaying the list.  
CC The polynucleotides are useful as probes, primers, sense and antisense  
CC nucleotides. The subtilisin homologue polypeptides are useful for  
CC producing antibodies which have diagnostic uses related to the activity,  
CC distribution and expression of subtilisin homologues. The subtilisin  
CC homologue polypeptides are used in compositions that serve as cleaning  
CC solutions in a wide variety of applications including laundry detergents,  
CC contact lens cleansing solutions, and dry cleaning. AAU38744-AAU38874  
CC represent the amino acid sequences of novel subtilisin homologues of the

CC invention.  
 XX Sequence 173 AA;  
 SQ

Query Match 91.7%; Score 22; DB 22; Length 173;  
 Best Local Similarity 66.7%; Pred. No. 6.3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 71 gtsps 76

RESULT 42  
 AAU38804  
 ID AAU38804 standard; Protein; 173 AA.  
 XX  
 AC AAU38804;  
 XX  
 DT 19-DEC-2001 (first entry)  
 XX  
 DE Novel subtilisin homologue #61.  
 XX  
 KW Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
 KW contact lens cleansing solution; dry cleaning.  
 XX  
 OS Bacillus sp.  
 XX  
 PN WO200175087-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 02-APR-2001; 2001WO-US10781.  
 XX  
 PR 03-APR-2000; 2000US-194143P.  
 XX  
 PA (MAXY-) MAXYGEN INC.  
 XX  
 PI Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
 PI Minshull J;  
 XX  
 XX WPI; 2001-616689/71.  
 DR N-PSDB; AAS56659.  
 XX  
 Novel subtilisin homologue polypeptides having improved endo-protease  
 activity relative to mature Savinase subtilisin polypeptide, useful as  
 component of cleaning solutions e.g. laundry detergents, dry cleaning

Claim 1; Page 109; 138pp; English.

The invention relates to an isolated polypeptide with improved endo-  
 protease activity relative to subtilisin homologue polypeptide, Savinase  
 (RTM). The invention also relates to an integrated system comprising a  
 computer or computer readable medium comprising a database comprising one  
 or more sequence records. Each record comprises one or more character  
 string corresponding to a nucleic acid or protein sequence of AAS56599-  
 AAS56728 or AAU38744-AAU38873 respectively. The integrated system  
 comprises a user input interface allowing a user to selectively  
 one or more sequence record. The integrated system is useful for  
 presenting information pertaining to one of several sequence records  
 stored in a database. The method involves determining a list of one or  
 more character strings corresponding to the sequence or its subsequence,  
 determining which character strings of the list are selected by a user,  
 and displaying the selected character strings or aligning the selected  
 character string with an additional character string. The method  
 further involves displaying an alignment of the selected character  
 string with the additional character string, and displaying the list.  
 The polynucleotides are useful as probes, primers, sense and antisense  
 nucleotides. The subtilisin homologue polypeptides are useful for  
 producing antibodies which have diagnostic uses related to the activity,  
 distribution and expression of subtilisin homologues. The subtilisin

CC homologue polypeptides are used in compositions that serve as cleaning  
 CC solutions in a wide variety of applications including laundry detergents,  
 CC contact lens cleansing solutions, and dry cleaning. AAU38744-AAU38874  
 CC represent the amino acid sequences of novel subtilisin homologues of the  
 CC invention.  
 XX  
 SQ Sequence 173 AA;  
 Query Match 91.7%; Score 22; DB 22; Length 173;  
 Best Local Similarity 66.7%; Pred. No. 6.3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 71 gtsaps 76

RESULT 43  
 AAU38810  
 ID AAU38810 standard; Protein; 173 AA.  
 XX  
 AC AAU38810;  
 XX  
 DT 19-DEC-2001 (first entry)  
 XX  
 DE Novel subtilisin homologue #67.  
 XX  
 KW Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
 KW contact lens cleansing solution; dry cleaning.  
 XX  
 OS Bacillus sp.  
 XX  
 PN WO200175087-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 02-APR-2001; 2001WO-US10781.  
 XX  
 PR 03-APR-2000; 2000US-194143P.  
 XX  
 PA (MAXY-) MAXYGEN INC.  
 XX  
 PI Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
 PI Minshull J;  
 XX  
 XX WPI; 2001-616689/71.  
 DR N-PSDB; AAS56665.  
 XX  
 Novel subtilisin homologue polypeptides having improved endo-protease  
 activity relative to mature Savinase subtilisin polypeptide, useful as  
 component of cleaning solutions e.g. laundry detergents, dry cleaning

Claim 1; Page 110; 138pp; English.

The invention relates to an isolated polypeptide with improved endo-  
 protease activity relative to subtilisin homologue polypeptide, Savinase  
 (RTM). The invention also relates to an integrated system comprising one  
 computer or computer readable medium comprising a database comprising one  
 or more sequence records. Each record comprises one or more character  
 string corresponding to a nucleic acid or protein sequence of AAS56599-  
 AAS56728 or AAU38744-AAU38873 respectively. The integrated system  
 comprises a user input interface allowing a user to selectively  
 one or more sequence record. The integrated system is useful for  
 presenting information pertaining to one of several sequence records  
 stored in a database. The method involves determining a list of one or  
 more character strings corresponding to the sequence or its subsequence,  
 determining which character strings of the list are selected by a user,  
 and displaying the selected character strings or aligning the selected  
 character string with an additional character string. The method  
 further involves displaying an alignment of the selected character  
 string with the additional character string, and displaying the list.  
 The polynucleotides are useful as probes, primers, sense and antisense  
 nucleotides. The subtilisin homologue polypeptides are useful for  
 producing antibodies which have diagnostic uses related to the activity,  
 distribution and expression of subtilisin homologues. The subtilisin

CC The polynucleotides are useful as probes, primers, sense and antisense  
 CC nucleotides. The subtilisin homologue polypeptides are useful for  
 CC producing antibodies which have diagnostic uses related to the activity,  
 CC distribution and expression of subtilisin homologues. The subtilisin  
 CC homologue polypeptides are used in compositions that serve as cleaning  
 CC solutions in a wide variety of applications including laundry detergents,  
 CC contact lens cleansing solutions, and dry cleaning. AAU38744-AAU38874  
 CC represent the amino acid sequences of novel subtilisin homologues of the  
 CC invention.

XX Sequence 173 AA;

Query Match 91.7%; Score 22; DB 22; Length 173;

Best Local Similarity 66.7%; Pred. No. 6.3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 DB 71 gtsaps 76

RESULT 44  
 AAU38835  
 ID AAU38835 standard; Protein; 173 AA.

XX AC AAU38835;

DT 19-DEC-2001 (first entry)

XX Novel subtilisin homologue #92.

XX Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
 KW contact lens cleansing solution; dry cleaning.

XX Bacillus sp.

XX WO200175087-A2.

XX 11-OCT-2001.

XX 02-APR-2001; 2001WO-US10781.

XX 03-APR-2000; 2000US-194143P.

XX (MAXY-) MAXYGEN INC.

XX Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
 PI Minshull J;

XX WPI; 2001-616689/71.

XX N-PSDB; AAS56690.

XX Novel subtilisin homologue polypeptides having improved endo-protease  
 PT activity relative to mature Savinase subtilisin polypeptide, useful as  
 PT component of cleaning solutions e.g. laundry detergents, dry cleaning

PS Claim 1; Page 113; 138pp; English.

XX The invention relates to an isolated polypeptide with improved endo-  
 CC protease activity relative to subtilisin homologue polypeptide, Savinase  
 CC (RTM). The invention also relates to an integrated system comprising a  
 CC computer or computer readable medium comprising a database comprising one  
 CC or more sequence records. Each record comprises one or more character  
 CC string corresponding to a nucleic acid or protein sequence of AAS56599-  
 CC AAS56728 or AAU38744-AAU38873 respectively. The integrated system  
 CC comprises a user input interface allowing a user to selectively  
 CC one or more sequence record. The integrated system is useful for  
 CC presenting information pertaining to one of several sequence records  
 CC stored in a database. The method involves determining a list of one or  
 CC more character strings corresponding to the sequence or its subsequence,  
 CC determining which character strings of the list are selected by a user,

CC and displaying the selected character strings or aligning the selected  
 CC character string with an additional character string. The method  
 CC further involves displaying an alignment of the selected character  
 CC string with the additional character string, and displaying the list.  
 CC The polynucleotides are useful as probes, primers, sense and antisense  
 CC nucleotides. The subtilisin homologue polypeptides are useful for  
 CC producing antibodies which have diagnostic uses related to the activity,  
 CC distribution and expression of subtilisin homologues. The subtilisin  
 CC homologue polypeptides are used in compositions that serve as cleaning  
 CC solutions in a wide variety of applications including laundry detergents,  
 CC contact lens cleansing solutions, and dry cleaning. AAU38744-AAU38874  
 CC represent the amino acid sequences of novel subtilisin homologues of the  
 CC invention.

XX Sequence 173 AA;

Query Match 91.7%; Score 22; DB 22; Length 173;

Best Local Similarity 66.7%; Pred. No. 6.3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 DB 71 gtsaps 76

RESULT 45

AAU38843  
 ID AAU38843 standard; Protein; 173 AA.

XX AC AAU38843;

DT 19-DEC-2001 (first entry)

XX Novel subtilisin homologue #100.

XX Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;  
 KW contact lens cleansing solution; dry cleaning.

XX Bacillus sp.

XX WO200175087-A2.

XX 11-OCT-2001.

XX 02-APR-2001; 2001WO-US10781.

XX 03-APR-2000; 2000US-194143P.

XX (MAXY-) MAXYGEN INC.

XX Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;  
 PI Minshull J;

XX WPI; 2001-616689/71.

XX N-PSDB; AAS56698.

XX Novel subtilisin homologue polypeptides having improved endo-protease  
 PT activity relative to mature Savinase subtilisin polypeptide, useful as  
 PT component of cleaning solutions e.g. laundry detergents, dry cleaning

PS Claim 1; Page 114; 138pp; English.

XX The invention relates to an isolated polypeptide with improved endo-  
 CC protease activity relative to subtilisin homologue polypeptide, Savinase  
 CC (RTM). The invention also relates to an integrated system comprising a  
 CC computer or computer readable medium comprising a database comprising one  
 CC or more sequence records. Each record comprises one or more character  
 CC string corresponding to a nucleic acid or protein sequence of AAS56599-  
 CC AAS56728 or AAU38744-AAU38873 respectively. The integrated system  
 CC comprises a user input interface allowing a user to selectively  
 CC one or more sequence record. The integrated system is useful for

CC presenting information pertaining to one of several sequence records  
 CC stored in a database. The method involves determining a list of one or  
 CC more character strings corresponding to the sequence or its subsequence,  
 CC determining which character strings of the list are selected by a user,  
 CC and displaying the selected character strings or aligning the selected  
 CC character string with an additional character string. The method  
 CC further involves displaying an alignment of the selected character  
 CC string with the additional character string, and displaying the list.  
 CC The polynucleotides are useful as probes, primers, sense and antisense  
 CC nucleotides. The subtilisin homologue polypeptides are useful for  
 CC producing antibodies which have diagnostic uses related to the activity,  
 CC distribution and expression of subtilisin homologues. The subtilisin  
 CC homologue polypeptides are used in compositions that serve as cleaning  
 CC solutions in a wide variety of applications including laundry detergents,  
 CC contact lens cleansing solutions, and dry cleaning. AAU38744-AAU38874  
 CC represent the amino acid sequences of novel subtilisin homologues of the  
 CC invention.

XX Sequence 173 AA;

Query Match 91.7%; Score 22; DB 22; Length 173;

Best Local Similarity 66.7%; Pred. No. 6.3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 71 gtspps 76

RESULT 46

AAU38869

ID AAU38869 standard; Protein; 173 AA.

AC AAU38869;

DT 19-DEC-2001 (first entry)

XX Novel subtilisin homologue #126.

DE Subtilisin; Savinase; diagnostic; cleaning solution; laundry detergent;

KW contact lens cleansing solution; dry cleaning.

XX Bacillus sp.

XX WO200175087-A2.

XX 11-OCT-2001.

XX 02-APR-2001; 2001WO-US10781.

XX 03-APR-2000; 2000US-194143P.

XX (MAXY-) MAXYGEN INC.

XX Ness JE, Welch M, Giver LJ, Cherry JR, Borchert TV, Stemmer WPC;

PI Minshull J;

XX WPI: 2001-616689/71.

DR N-PSDB; AAS56724.

XX Novel subtilisin homologue polypeptides having improved endo-protease

PT activity relative to mature Savinase subtilisin polypeptide, useful as

PT component of cleaning solutions e.g. laundry detergents, dry cleaning

PT .

XX Claim 1; Page 117; 138pp; English.

XX The invention relates to an isolated polypeptide with improved endo-

CC protease activity relative to subtilisin homologue polypeptide, Savinase

CC (RPM). The invention also relates to an integrated system comprising a

CC computer or computer readable medium comprising a database comprising one

CC or more sequence records. Each record comprises one or more character

CC string corresponding to a nucleic acid or protein sequence of AAS56599-  
 CC AAS56728 or AAU38744-AAU38873 respectively. The integrated system  
 CC comprises a user input interface allowing a user to selectively  
 CC one or more sequence record. The integrated system is useful for  
 CC presenting information pertaining to one of several sequence records  
 CC stored in a database. The method involves determining a list of one or  
 CC more character strings corresponding to the sequence or its subsequence,  
 CC determining which character strings of the list are selected by a user,  
 CC and displaying the selected character strings or aligning the selected  
 CC character string with an additional character string. The method  
 CC further involves displaying an alignment of the selected character  
 CC string with the additional character string, and displaying the list.  
 CC The polynucleotides are useful as probes, primers, sense and antisense  
 CC nucleotides. The subtilisin homologue polypeptides are useful for  
 CC producing antibodies which have diagnostic uses related to the activity,  
 CC distribution and expression of subtilisin homologues. The subtilisin  
 CC homologue polypeptides are used in compositions that serve as cleaning  
 CC solutions in a wide variety of applications including laundry detergents,  
 CC contact lens cleansing solutions, and dry cleaning. AAU38744-AAU38874  
 CC represent the amino acid sequences of novel subtilisin homologues of the  
 CC invention.

SQ Sequence 173 AA;

Query Match 91.7%; Score 22; DB 22; Length 173;

Best Local Similarity 66.7%; Pred. No. 6.3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 71 gtspps 76

RESULT 47

ABB68003

ID ABB68003 standard; Protein; 175 AA.

XX ABB68003;

AC 26-MAR-2002 (first entry)

DT Drosophila melanogaster polypeptide SEQ ID NO 30801.

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI: 2001-656860/75.

DR N-PSDB; ABL12106.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions .

XX Disclosure; SEQ ID NO 30801; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 175 AA;

Query Match 91.7%; Score 22; DB 22; Length 175;  
Best Local Similarity 66.7%; Pred. No. 6.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
DB 98 gdstps 103  
|| ||

RESULT 48  
ABB68002  
ID ABB68002 standard; Protein; 182 AA.  
XX  
AC ABB68002;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 30798.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
XX N-PSDB; ABL12105.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 30798; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 182 AA;

Query Match 91.7%; Score 22; DB 22; Length 182;  
Best Local Similarity 66.7%; Pred. No. 6.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
DB 105 gdstps 110  
|| ||

RESULT 49  
AAM40840  
ID AAM40840 standard; Protein; 190 AA.  
XX  
AC AAM40840;  
XX  
DT 22-OCT-2001 (first entry)  
XX  
XX Human polypeptide SEQ ID NO 5771.  
XX  
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia.  
XX  
XX Homo sapiens.  
OS  
XX WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PF 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI: 2001-442253/47.  
XX N-PSDB; AAI59996.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
XX Example 2; SEQ ID NO 5771; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed

```
CC specification.
SQ Sequence 190 AA;

Query Match          91.7%; Score 22; DB 22; Length 190;
Best Local Similarity 66.7%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
   || ||
Db 39 gtstps 44

RESULT 50
AAU40065
ID AAU40065 standard; Protein; 194 AA.
XX
AC AAU40065;
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #961.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD
XX
11-NOV-2001.
XX
20-APR-2001; 2001WO-US12865.
XX
21-APR-2000; 2000US-199047P.
XX
02-JUN-2000; 2000US-208841P.
XX
07-JUL-2000; 2000US-216747P.
XX
(CORI-) CORIXA CORP.
XX
Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
WPI; 2001-616774/71.
DR N-PSDB; AAS59510.
XX
Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
Example 1; SEQ ID No 1260; 1069pp; English.
XX
Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC
Note: The sequence data for this patent did not form part of the printed
```

```
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 194 AA;

Query Match          91.7%; Score 22; DB 22; Length 194;
Best Local Similarity 66.7%; Pred. No. 7e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
   || ||
Db 48 gtssps 53

Search completed: September 3, 2002, 09:26:28
Job time: 80 sec
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 09:25:08 ; Search time 12.98 Seconds  
(without alignments)  
11.291 Million cell updates/sec

Title: BASK-853-CLAIM5  
Perfect score: 24  
Sequence: 1 gtxxps 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTCUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	91.7	38	1 US-08-176-500-41	Sequence 41, Appl
2	22	91.7	38	1 US-08-471-052A-41	Sequence 41, Appl
3	22	91.7	38	1 US-08-189-331-41	Sequence 41, Appl
4	22	91.7	38	2 US-08-471-939-41	Sequence 41, Appl
5	22	91.7	38	2 US-08-471-800-41	Sequence 41, Appl
6	22	91.7	38	2 US-08-471-068-41	Sequence 41, Appl
7	22	91.7	166	1 US-08-665-617-2	Sequence 2, Appl
8	22	91.7	168	2 US-08-353-476-73	Sequence 73, Appl
9	22	91.7	241	2 US-08-353-476-116	Sequence 116, App
10	22	91.7	273	2 US-08-353-476-112	Sequence 112, App
11	22	91.7	294	4 US-09-372-422A-40	Sequence 40, Appl
12	22	91.7	1050	4 US-09-428-711A-16	Sequence 16, Appl
13	21	87.5	40	5 PCT-US96-08730-14	Sequence 14, Appl
14	21	87.5	56	2 US-08-537-400-32	Sequence 32, Appl
15	21	87.5	153	4 US-09-228-986-83	Sequence 83, Appl
16	21	87.5	170	3 US-08-483-101-3	Sequence 3, Appl
17	21	87.5	170	3 US-09-081-180-5	Sequence 5, Appl
18	21	87.5	170	3 US-09-040-786-5	Sequence 5, Appl
19	21	87.5	225	1 US-08-290-979A-8	Sequence 8, Appl
20	21	87.5	295	2 US-08-481-956A-9	Sequence 9, Appl
21	21	87.5	295	2 US-08-629-291A-9	Sequence 9, Appl
22	21	87.5	295	2 US-08-658-335B-9	Sequence 9, Appl
23	21	87.5	334	2 US-08-997-080-45	Sequence 45, Appl
24	21	87.5	334	2 US-08-997-362-45	Sequence 45, Appl
25	21	87.5	334	3 US-08-873-970-45	Sequence 45, Appl
26	21	87.5	334	4 US-09-095-855-45	Sequence 45, Appl
27	21	87.5	334	4 US-08-705-347A-45	Sequence 45, Appl

28	21	87.5	334	4 US-09-324-542-45	Sequence 45, Appl
29	21	87.5	335	4 US-09-198-955A-10	Sequence 10, Appl
30	21	87.5	335	4 US-09-184-217-1	Sequence 1, Appl
31	21	87.5	340	4 US-09-188-930-184	Sequence 184, App
32	21	87.5	385	3 US-09-053-866-2	Sequence 2, Appl
33	21	87.5	389	1 US-08-409-199-3	Sequence 3, Appl
34	21	87.5	439	4 US-09-171-461-32	Sequence 32, Appl
35	21	87.5	446	3 US-08-956-254-2	Sequence 2, Appl
36	21	87.5	446	3 US-09-008-388-1	Sequence 1, Appl
37	21	87.5	448	2 US-09-015-815-1	Sequence 1, Appl
38	21	87.5	448	3 US-08-476-509B-2	Sequence 2, Appl
39	21	87.5	486	3 US-08-348-518C-2	Sequence 2, Appl
40	21	87.5	495	2 US-08-794-795-2	Sequence 2, Appl
41	21	87.5	495	4 US-09-249-200-2	Sequence 2, Appl
42	21	87.5	534	2 US-09-031-392-4	Sequence 4, Appl
43	21	87.5	534	4 US-09-299-549-4	Sequence 4, Appl
44	21	87.5	534	4 US-09-610-417-4	Sequence 4, Appl
45	21	87.5	577	4 US-09-413-814-4	Sequence 4, Appl
46	21	87.5	732	2 US-08-843-530B-28	Sequence 28, Appl
47	21	87.5	739	2 US-08-836-943-2	Sequence 2, Appl
48	21	87.5	774	2 US-08-231-193A-42	Sequence 42, Appl
49	21	87.5	774	2 US-08-486-273A-42	Sequence 42, Appl
50	21	87.5	774	3 US-08-480-474-42	Sequence 42, Appl

ALIGNMENTS

RESULT 1  
US-08-176-500-41  
; Sequence 41, Application US/08176500  
; Patent No. 5498538  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.  
; APPLICANT: Fowlkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/176,500  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/013,416  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mirock, S. Leslie  
; REGISTRATION NUMBER: 18,972  
; REFERENCE/DOCKET NUMBER: 1101-143  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 38 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-176-500-41

Query Match 91.7%; Score 22; DB 1; Length 38;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtxxps 6  
|| ||  
Db 20 GTTSPS 25

RESULT 2  
US-08-471-052A-41  
; Sequence 41, Application US/08471052A  
; Patent No. 5625033  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.  
; APPLICANT: Fowlkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 166  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,052A  
; FILING DATE: 06-JUNE-1995  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-179  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 38 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-471-052A-41

Query Match 91.7%; Score 22; DB 1; Length 38;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtxxps 6  
|| ||  
Db 20 GTTSPS 25

RESULT 3  
US-08-189-331-41  
; Sequence 41, Application US/08189331  
; Patent No. 5747334  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.  
; APPLICANT: Fowlkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 186  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/189,331  
; FILING DATE: Concurrently herewith  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-155  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212 790-9090  
; TELEFAX: 212 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 38 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
US-08-189-331-41

Query Match 91.7%; Score 22; DB 1; Length 38;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtxxps 6  
|| ||  
Db 20 GTTSPS 25

RESULT 4  
US-08-471-939-41  
; Sequence 41, Application US/08471939  
; Patent No. 5844076  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.  
; APPLICANT: Fowlkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,939  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/013,416  
; FILING DATE: 01-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-143

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-471-939-41

Query Match 91.7%; Score 22; DB 2; Length 38;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 20 GTTSPS 25

RESULT 5  
US-08-471-800-41  
; Sequence 41, Application US/08471800  
; Patent No. 5852167  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.

; APPLICANT: Fowkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 141  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,800  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/013,416  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-143  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-471-800-41

Query Match 91.7%; Score 22; DB 2; Length 38;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 20 GTTSPS 25

RESULT 6  
US-08-471-068-41  
; Sequence 41, Application US/08471068  
; Patent No. 5948635  
; GENERAL INFORMATION:  
; APPLICANT: Kay, B. K.

; APPLICANT: Fowkes, D. M.  
; TITLE OF INVENTION: Totally Synthetic Affinity Reagents  
; NUMBER OF SEQUENCES: 186  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471,068  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/189,331  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-155  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212 790-9090  
TELEFAX: 212 869-8864/9741  
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 38 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-471-068-41

Query Match 91.7%; Score 22; DB 2; Length 38;  
Best Local Similarity 66.7%; Pred. No. 68;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 20 GTTSPS 25

RESULT 7  
US-08-665-617-2  
; Sequence 2, Application US/08665617  
; Patent No. 5663316  
; GENERAL INFORMATION:  
; APPLICANT: Xudong, Yin

; TITLE OF INVENTION: Gene and Protein for Regulation of Cell Death  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st Street, Suite A-1  
; CITY: Gainesville  
; STATE: Florida

US-08-665-617-2

; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/665,617  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saliwanchik, David R.  
; REGISTRATION NUMBER: 31,794  
; REFERENCE/DOCKET NUMBER: CL-8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (352) 375-8100  
; TELEFAX: (352) 372-5800  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 166 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-665-617-2

Query Match 91.7%; Score 22; DB 1; Length 166;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|||  
Db 160 GTAAPS 165

RESULT 8  
US-08-353-476-73  
; Sequence 73, Application US/08353476  
; Patent No. 5871902  
; GENERAL INFORMATION:  
; APPLICANT: Weininger, Susan  
; TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A  
; TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st St., Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,476  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bencen, Gerard H  
; REGISTRATION NUMBER: 35,746  
; REFERENCE/DOCKET NUMBER: GP-100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (904) 375-8100  
; TELEFAX: (904) 372-5800  
; INFORMATION FOR SEQ ID NO: 73:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 168 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
US-08-353-476-73

Query Match 91.7%; Score 22; DB 2; Length 168;  
Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|||  
Db 119 GTATPS 124

RESULT 9  
US-08-353-476-116  
; Sequence 116, Application US/08353476  
; Patent No. 5871902  
; GENERAL INFORMATION:  
; APPLICANT: Weininger, Susan  
; APPLICANT: Weininger, Arthur M  
; TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A  
; TITLE OF INVENTION: SPECIFIC SEQUENCE COMPOSITION  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st St., Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,476  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bencen, Gerard H  
; REGISTRATION NUMBER: 35,746  
; REFERENCE/DOCKET NUMBER: GP-100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (904) 375-8100  
; TELEFAX: (904) 372-5800  
; INFORMATION FOR SEQ ID NO: 116:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 241 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-08-353-476-116

Query Match 91.7%; Score 22; DB 2; Length 241;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|||  
Db 192 GTATPS 197

RESULT 10

US-08-353-476-112  
; Sequence 112, Application US/08353476  
; Patent No. 5871902  
; GENERAL INFORMATION:  
; APPLICANT: Weininger, Susan  
; APPLICANT: Weininger, Arthur M  
; TITLE OF INVENTION: METHOD OF DETECTION OF DNA WITH A  
; SPECIFIC SEQUENCE COMPOSITION  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Saliwanchik & Saliwanchik  
; STREET: 2421 N.W. 41st St., Suite A-1  
; CITY: Gainesville  
; STATE: Florida  
; COUNTRY: USA  
; ZIP: 32606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/353,476  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bencen, Gerard H  
; REGISTRATION NUMBER: 35,746  
; REFERENCE/DOCKET NUMBER: GP-100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (904) 375-8100  
; TELEFAX: (904) 372-5800  
; INFORMATION FOR SEQ ID NO: 112:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 273 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-08-353-476-112

Query Match 91.7%; Score 22; DB 2; Length 273;  
Best Local Similarity 66.7%; Pred. No. 4.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 224 GTATPS 229

RESULT 11  
US-09-372-422A-40  
; Sequence 40, Application US/09372422A  
; Patent No. 6313375  
; GENERAL INFORMATION:  
; APPLICANT: Rudolf Jung  
; APPLICANT: Francois Barrieu  
; TITLE OF INVENTION: Maize Aquaporins and Uses Thereof  
; FILE REFERENCE: 0919  
; CURRENT APPLICATION NUMBER: US/09/372,422A  
; CURRENT FILING DATE: 1999-08-11  
; PRIOR APPLICATION NUMBER: US 60/098,692  
; PRIOR FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 40  
; LENGTH: 294  
; TYPE: PRT  
; ORGANISM: zea mays  
US-09-372-422A-40

Query Match 91.7%; Score 22; DB 4; Length 294;  
Best Local Similarity 66.7%; Pred. No. 4.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gtxxps 6  
|| ||  
Db 158 GTTTPS 163  
RESULT 12  
US-09-428-711A-16  
; Sequence 16, Application US/09428711A  
; Patent No. 6358720  
; GENERAL INFORMATION:  
; APPLICANT: Muramatsu, Masaaki  
; APPLICANT: Shirasawa, Takuji  
; APPLICANT: Tokumitsu, Hiroshi  
; APPLICANT: No. 6358720uchi, Teruhisa  
; TITLE OF INVENTION: SERINE/THREONINE PROTEIN KINASE  
; FILE REFERENCE: 06501-045001  
; CURRENT APPLICATION NUMBER: US/09/428,711A  
; CURRENT FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: PCT/JP98/01246  
; PRIOR FILING DATE: 1998-03-23  
; PRIOR APPLICATION NUMBER: JP 9/124798  
; PRIOR FILING DATE: 1997-04-28  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 1050  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-428-711A-16

Query Match 91.7%; Score 22; DB 4; Length 1050;  
Best Local Similarity 66.7%; Pred. No. 1.4e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 745 GTSSPS 750

RESULT 13  
PCT-US96-08730-14  
; Sequence 14, Application PC/TUS9608730  
; GENERAL INFORMATION:  
; APPLICANT: Cassels, Frederick  
; APPLICANT: Anderson, Jeffrey  
; APPLICANT: Carter, John Mark  
; TITLE OF INVENTION: Methods of Raising Antibodies Against E.  
; COLI OF THE FAMILY CSF-CFA /1  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Glenna Hendricks  
; STREET: P.O. Box 2509  
; CITY: Fairfax  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22031  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/08730  
; FILING DATE: 03-JUN-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:

NAME: Hendricks, Glenna  
REGISTRATION NUMBER: 32,535  
REFERENCE/DOCKET NUMBER: PCT/US96/08730  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 591-4470  
TELEFAX: (703) 591-4428  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 40 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
PCT-US96-08730-14

Query Match 87.5%; Score 21; DB 5; Length 40;  
Best Local Similarity 66.7%; Pred. No. 1.2e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 22 GTALPS 27

RESULT 14  
US-08-537-400-32  
Sequence 32, Application US/08537400  
Patent No. 5939301  
GENERAL INFORMATION:  
APPLICANT: Cloned DNA Polymerases From Thermotoga  
TITLE OF INVENTION: neapolitana And Mutants Thereof  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.  
STREET: 1100 New York Avenue, N.W., Suite 600  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/537,400  
FILING DATE: 02-OCT-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/316,423  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/370,190  
FILING DATE: 09-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0942.2800002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-2600  
TELEFAX: 202-371-2540  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-08-537-400-32

Query Match 87.5%; Score 21; DB 2; Length 56;  
Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 44 GTRTPS 49

RESULT 15  
US-09-228-986-83  
Sequence 83, Application US/09228986  
Patent No. 6359198  
GENERAL INFORMATION:  
APPLICANT: Strabala, Timothy  
TITLE OF INVENTION: Compositions Isolated from Plant Cells  
TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signalling  
FILE REFERENCE: 11000/1020  
CURRENT APPLICATION NUMBER: US/09/228,986  
CURRENT FILING DATE: 1999-01-12  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: FastSEQ for Windows Version 3.0  
SEQ ID NO 83  
LENGTH: 153  
TYPE: PRT  
ORGANISM: Pinus radiata  
US-09-228-986-83

Query Match 87.5%; Score 21; DB 4; Length 153;  
Best Local Similarity 66.7%; Pred. No. 4.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 103 GTLSPS 108

RESULT 16  
US-08-483-101-3  
Sequence 3, Application US/08483101  
Patent No. 5932715  
GENERAL INFORMATION:  
APPLICANT: Scott, June R.  
APPLICANT: Froehlich, Barbara  
APPLICANT: Caron, Judy  
TITLE OF INVENTION: CS2 Proteins and Coding Sequences  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee and Winner, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,101  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ferber, Donna M.  
REGISTRATION NUMBER: 33878  
REFERENCE/DOCKET NUMBER: 6-95  
TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 499-8080  
; TELEFAX: (303) 499-8089  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 170 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-483-101-3

Query Match 87.5%; Score 21; DB 2; Length 170;  
Best Local Similarity 66.7%; Pred. No. 4.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 45 GTALPS 50

RESULT 17  
US-09-081-180-5  
; Sequence 5, Application US/09081180  
; Patent No. 6022847  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG32  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics  
; STREET: 1201 Eastlake Ave. E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081,180  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/041,263  
; FILING DATE: March 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lingenfelter, Susan E  
; REGISTRATION NUMBER: 41,156  
; REFERENCE/DOCKET NUMBER: 97-17C1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6675  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 170 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-081-180-5

Query Match 87.5%; Score 21; DB 3; Length 170;  
Best Local Similarity 66.7%; Pred. No. 4.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||

Db 95 GTYSPS 100

RESULT 18  
US-09-040-786-5  
; Sequence 5, Application US/09040786  
; Patent No. 6025197  
; GENERAL INFORMATION:  
; APPLICANT: Sheppard, Paul O.  
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG32  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ZymoGenetics  
; STREET: 1201 Eastlake Ave. E.  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/040,786  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/041,263  
; FILING DATE: March 19, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lingenfelter, Susan E  
; REGISTRATION NUMBER: 41,156  
; REFERENCE/DOCKET NUMBER: 97-17  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6675  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 170 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-040-786-5

Query Match 87.5%; Score 21; DB 3; Length 170;  
Best Local Similarity 66.7%; Pred. No. 4.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 95 GTYSPS 100

RESULT 19  
US-08-290-979A-8  
; Sequence 8, Application US/08290979A  
; Patent No. 5610046  
; GENERAL INFORMATION:  
; APPLICANT: VAN OOLJEN, Albert J.H.  
; APPLICANT: DE GRAFF, Leendert H.  
; APPLICANT: VAN DEN BROECK, Henriette C.  
; APPLICANT: VISSER, Jacob  
; TITLE OF INVENTION: Cloning and Expression of Xylanase B  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W.

; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20006-1812  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,979A  
; FILING DATE: 22-SEP-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KATE H. MURASHIGE  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 4615-0045.00  
; TELEPHONE: (202) 887-1500  
; TELEFAX: (202) 887-0763  
; TELEX: 90-4030 MRSN FOERS WSH  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 225 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-290-979A-8

Query Match 87.5%; Score 21; DB 1; Length 225;  
Best Local Similarity 66.7%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 99 GTFTPS 104

RESULT 20  
US-08-481-956A-9  
; Sequence 9, Application US/08481956A  
; Patent No. 5824867  
; GENERAL INFORMATION:  
; APPLICANT: Coruzzi, Gloria  
; APPLICANT: Oliveira, Igor  
; APPLICANT: Lam, Hon-Ming  
; APPLICANT: Hsieh, Ming-Hsiun  
; TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,956A  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 5914-040  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864

; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-481-956A-9

Query Match 87.5%; Score 21; DB 2; Length 295;  
Best Local Similarity 66.7%; Pred. No. 7.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 65 GTVSPS 70

RESULT 21  
US-08-629-291A-9  
; Sequence 9, Application US/08629291A  
; Patent No. 5959174  
; GENERAL INFORMATION:  
; APPLICANT: Coruzzi, Gloria  
; APPLICANT: Oliveira, Igor  
; APPLICANT: Lam, Hon-Ming  
; APPLICANT: Hsieh, Ming-Hsiun  
; TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/629,291A  
; FILING DATE: 08-APR-1996  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 5914-050  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 295 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-629-291A-9

Query Match 87.5%; Score 21; DB 2; Length 295;  
Best Local Similarity 66.7%; Pred. No. 7.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 65 GTVSPS 70

```
RESULT 22
US-08-658-335B-9
; Sequence 9, Application US/08658335B
; Patent No. 5981703
; GENERAL INFORMATION:
; APPLICANT: Coruzzi, Gloria
; APPLICANT: Oliveira, Igor
; APPLICANT: Lam, Hon-Ming
; APPLICANT: Hsieh, Ming-Hsiun
; TITLE OF INVENTION: PLANT GLUTAMATE RECEPTORS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,335B
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 5914-052
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 295 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-658-335B-9

Query Match 87.5%; Score 21; DB 2; Length 295;
Best Local Similarity 66.7%; Pred. No. 7.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
Db 65 GTVSPS 70

RESULT 23
US-08-997-080-45
; Sequence 45, Application US/08997080
; Patent No. 598524
; GENERAL INFORMATION:
; APPLICANT: WATSON, JAMES D.
; APPLICANT: TAN, PAUL L.J.
; TITLE OF INVENTION: METHODS AND COMPOUNDS FOR THE TREATMENT OF IMMUNOLOGICALLY-
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,080
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-997-080-45

Query Match 87.5%; Score 21; DB 2; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
Db 253 GTCTPS 258

RESULT 24
US-08-997-362-45
; Sequence 45, Application US/08997362
; Patent No. 5985287
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyma, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 194
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/997,362
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/873,970
; FILING DATE: June 12, 1997
; APPLICATION NUMBER: U.S. Patent Application No. 5985287 08/705,347
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-997-362-45

Query Match      87.5%; Score 21; DB 2; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
   || ||
Db 253 GTGTPS 258

RESULT 25
US-08-873-970-45
; Sequence 45, Application US/08873970
; Patent No. 6001361
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiayama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR
; TREATMENT AND DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,970
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-997-362-45

Query Match      87.5%; Score 21; DB 4; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
   || ||
Db 253 GTGTPS 258

; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-997-362-45

Query Match      87.5%; Score 21; DB 3; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
   || ||
Db 253 GTGTPS 258

; MOLECULE TYPE: protein
; US-08-873-970-45

Query Match      87.5%; Score 21; DB 3; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
   || ||
Db 253 GTGTPS 258

RESULT 26
US-09-095-855-45
; Sequence 45, Application US/09095855
; Patent No. 6160093
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Compounds and Methods for
; TREATMENT AND DIAGNOSIS OF Mycobacterial Infections
; NUMBER OF SEQUENCES: 208
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/095,855
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/705,347
; FILING DATE: 29-AUG-1996
; APPLICATION NUMBER: 08/873,970
; FILING DATE: 12-JUN-1997
; APPLICATION NUMBER: 08/997,362
; FILING DATE: 23-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002c3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-269-0565
; TELEFAX: 206-269-0563
; TELEX:
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-095-855-45

Query Match      87.5%; Score 21; DB 4; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
   || ||
Db 253 GTGTPS 258
```

```

; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-324-542-45

Query Match      87.5%; Score 21; DB 4; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 253 GTGTPS 258

RESULT 29
US-09-198-955A-10
; Sequence 10, Application US/09198955A
; Patent No. 6187580
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; FILE OF INVENTION: NO. 6187580el pectate Lyases
; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-198-955A-10

Query Match      87.5%; Score 21; DB 4; Length 335;
Best Local Similarity 66.7%; Pred. No. 8.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 84 GTITPS 89

RESULT 30
US-09-184-217-1
; Sequence 1, Application US/09184217
; Patent No. 6258590
; GENERAL INFORMATION:
; APPLICANT: Lange, Niels E.K.
; APPLICANT: Kongsbak, Lars
; APPLICANT: Schulein, Martin
; APPLICANT: Schulein, Martin

; Sequence 45, Application US/08705347A
; Patent No. 6284255
; GENERAL INFORMATION:
; APPLICANT: Tan, Paul
; APPLICANT: Hiyama, Jun
; APPLICANT: Visser, Elizabeth
; APPLICANT: Skinner, Margot
; APPLICANT: Scott, Linda
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF MYCOBACTERIAL INFECTIONS
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Speckman Picard PLLC
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/705,347A
; FILING DATE: 28-AUG-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sleath, Janet
; REGISTRATION NUMBER: 37,007
; REFERENCE/DOCKET NUMBER: 11000.1002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206.269.0565
; TELEFAX: 206.269.0563
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-705-347A-45

Query Match      87.5%; Score 21; DB 4; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 253 GTGTPS 258

RESULT 28
US-09-324-542-45
; Sequence 45, Application US/09324542
; Patent No. 6328978
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c1
; CURRENT APPLICATION NUMBER: US/09/324,542
; CURRENT FILING DATE: 1999-06-02
; EARLIER APPLICATION NUMBER: US 08/997,080
; EARLIER FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
```

; APPLICANT: Bjornvad, Mads E.  
; APPLICANT: Husain, Philip A.  
; TITLE OF INVENTION: Biopreparation Of Textiles At High  
; FILE OF INVENTION: Temperatures  
; FILE REFERENCE: 5729.000-US  
; CURRENT APPLICATION NUMBER: US/09/184,217  
; CURRENT FILING DATE: 1998-11-02  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 335  
; TYPE: PRT  
; ORGANISM: bacillus sp.  
US-09-184-217-1

Query Match 87.5%; Score 21; DB 4; Length 335;  
Best Local Similarity 66.7%; Pred. No. 8.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 84 GTTPS 89

RESULT 31  
US-09-188-930-184  
; Sequence 184, Application US/09188930A  
; Patent No. 6150502  
; GENERAL INFORMATION:  
; APPLICANT: Watson, James D.  
; APPLICANT: Strachan, Lorna  
; APPLICANT: Sleeman, Matthew  
; APPLICANT: Onrust, Rene  
; APPLICANT: Murrison, James Greg  
; TITLE OF INVENTION: Compositions Isolated From Skin Cells  
; FILE OF INVENTION: and Methods For Their Use  
; FILE REFERENCE: 11000.1011c1  
; CURRENT APPLICATION NUMBER: US/09/188,930A  
; CURRENT FILING DATE: 1998-11-09  
; NUMBER OF SEQ ID NOS: 348  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 184  
; LENGTH: 340  
; TYPE: PRT  
; ORGANISM: mouse  
US-09-188-930-184

Query Match 87.5%; Score 21; DB 4; Length 340;  
Best Local Similarity 66.7%; Pred. No. 8.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 120 STAGPS 125

RESULT 32  
US-09-053-866-2  
; Sequence 2, Application US/09053866  
; Patent No. 611075  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Wenfeng  
; APPLICANT: Presnell, Scott R.  
; APPLICANT: Yee, David P.  
; APPLICANT: Foster, Donald C.  
; TITLE OF INVENTION: PROTEASE-ACTIVATED RECEPTOR  
; FILE OF INVENTION: PAR4 (ZCHEMR2)  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Zymogenetics, Inc.  
; STREET: 1201 Eastlake Avenue East

; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/053,866  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Leith, Debra K  
; REGISTRATION NUMBER: 32,619  
; REFERENCE/DOCKET NUMBER: 98-10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6674  
; TELEFAX: 206-442-6678  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 385 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: Single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-09-053-866-2

Query Match 87.5%; Score 21; DB 3; Length 385;  
Best Local Similarity 66.7%; Pred. No. 9.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 19 GTQTPS 24

RESULT 33  
US-08-409-199-3  
; Sequence 3, Application US/08409199  
; Patent No. 5532153  
; GENERAL INFORMATION:  
; APPLICANT: XU, Shuang-yong  
; APPLICANT: XIAO, Jianping  
; TITLE OF INVENTION: METHOD FOR CLONING AND  
; TITLE OF INVENTION: PRODUCING THE SACL RESTRICTION  
; TITLE OF INVENTION: ENDONUCLEASE  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.  
; STREET: 32 TOZER ROAD  
; CITY: BEVERLY  
; STATE: MASSACHUSETTS  
; COUNTRY: US  
; ZIP: 01915  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/409,199  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILLIAMS, GREGORY D.

; REGISTRATION NUMBER: 30901  
; REFERENCE/DOCKET NUMBER: NEB-110  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (508) 927-5054; 292  
; TELEFAX: (508) 927-1705  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 389 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-409-199-3

Query Match 87.5%; Score 21; DB 1; Length 389;  
Best Local Similarity 66.7%; Pred. No. 9.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 375 GTTIPS 380

## RESULT 34

US-09-171-461-32  
; Sequence 32, Application US/09171461  
; Patent No. 6335016  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Adam  
; APPLICANT: Cotten, Matthew  
; APPLICANT: Chiocca, Susanna  
; APPLICANT: Kurzbaumer, Robert  
; APPLICANT: Schafner, Gotthold  
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus  
; FILE REFERENCE: 0652.180000  
; CURRENT APPLICATION NUMBER: US/09/171,461  
; EARLIER FILING DATE: 1999-01-12  
; EARLIER APPLICATION NUMBER: PCT/EP97/01944  
; EARLIER FILING DATE: 1997-04-18  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32  
; LENGTH: 439  
; TYPE: PRT  
; ORGANISM: CELO VIRUS  
; FEATURE:  
; OTHER INFORMATION: Position: 5366..6685/Product:IVa2  
US-09-171-461-32

Query Match 87.5%; Score 21; DB 4; Length 439;  
Best Local Similarity 66.7%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 19 GTKTPS 24

## RESULT 35

US-08-956-254-2  
; Sequence 2, Application US/08956254A  
; Patent No. 6013265  
; GENERAL INFORMATION:  
; APPLICANT: AURELIAN, LAURE  
; TITLE OF INVENTION: Vaccine Composition for Herpes Simplex Virus and  
; TITLE OF INVENTION: Methods of Using  
; FILE REFERENCE: 14211A  
; CURRENT APPLICATION NUMBER: US/08/956,254A  
; EARLIER FILING DATE: 1997-10-22  
; EARLIER APPLICATION NUMBER: US 60/029,093  
; EARLIER FILING DATE: 1996-10-22  
; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: herpes simplex virus-2  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1)..(446)  
; OTHER INFORMATION: PROTEIN KINASE DOMAIN OF ICP10 SUBUNIT OF HSV-2  
; PUBLICATION INFORMATION:  
; AUTHORS: CHUNG ET AL.,  
; JOURNAL: J. Virol.  
; VOLUME: 63  
; PAGES: 3389-3398  
; DATE: 1989  
; PUBLICATION INFORMATION:  
; AUTHORS: NELSON ET AL.,  
; JOURNAL: J. Biol. Chem.  
; VOLUME: 271  
; PAGES: 17021-17027  
; DATE: 1996  
US-08-956-254-2

Query Match 87.5%; Score 21; DB 3; Length 446;  
Best Local Similarity 66.7%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 116 GTSGPS 121

## RESULT 36

US-09-008-388-1  
; Sequence 1, Application US/09008388  
; Patent No. 6054131  
; GENERAL INFORMATION:  
; APPLICANT: AURELIAN, LAURE  
; TITLE OF INVENTION: VACCINE COMPOSITION FOR HERPES  
; TITLE OF INVENTION: SIMPLEX VIRUS AND METHOD OF USING  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PEPPER HAMILTON LLP  
; STREET: 600 FOURTEENTH STREET, N.W., Suite 500  
; CITY: WASHINGTON  
; STATE: DC  
; COUNTRY: U.S.A.  
; ZIP: 20005-2004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-WINDOWS  
; SOFTWARE: WP 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/008,388  
; FILING DATE: December 2, 1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ramsey, William S.  
; REGISTRATION NUMBER: 32,715  
; REFERENCE/DOCKET NUMBER: 114178.100  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 220-1280  
; TELEFAX: (202) 220-1665  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 446  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO

FRAGMENT TYPE: INTERNAL  
ORIGINAL SOURCE:  
ORGANISM: HERPES SIMPLEX  
STRAIN: RECOMBINANT  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
US-09-008-388-1

Query Match 87.5%; Score 21; DB 3; Length 446;  
Best Local Similarity 66.7%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|||  
Db 116 GTSGPS 121

RESULT 37  
US-09-015-815-1  
; Sequence 1, Application US/09015815  
; Patent No. 5965356  
; GENERAL INFORMATION:  
; APPLICANT: AURELIAN, LAURE  
; APPLICANT: SMITH, CYNTHIA  
; TITLE OF INVENTION: HERPES SIMPLEX VIRUS TYPE SPECIFIC SEROASSAY  
; FILE REFERENCE: 1437LA  
; CURRENT APPLICATION NUMBER: US/09/015,815  
; CURRENT FILING DATE: 1998-01-29  
; EARLIER APPLICATION NUMBER: US 60/036,622  
; EARLIER FILING DATE: 1997-01-31  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 448  
; TYPE: PRT  
; ORGANISM: herpes simplex virus-2  
; FEATURE:  
; NAME/KEY: DOMAIN  
; LOCATION: (1)..(446)  
; OTHER INFORMATION: PROTEIN KINASE DOMAIN OF ICP10 SUBUNIT OF HSV-2  
; PUBLICATION INFORMATION:  
; AUTHORS: CHUNG ET AL.,  
; JOURNAL: J. Virol.,  
; VOLUME: 63  
; PAGES: 3389-3398  
; DATE: 1989  
; PUBLICATION INFORMATION:  
; AUTHORS: NELSON ET AL.,  
; JOURNAL: J. Biol. Chem.,  
; VOLUME: 271  
; PAGES: 17021-17027  
; DATE: 1996  
US-09-015-815-1

Query Match 87.5%; Score 21; DB 2; Length 448;  
Best Local Similarity 66.7%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|||  
Db 116 GTSGPS 121

RESULT 38  
US-08-476-509B-2  
; Sequence 2, Application US/08476509B  
; Patent No. 6034212  
; GENERAL INFORMATION:  
; APPLICANT: SUDOL, MARIUS  
; APPLICANT: PEER, BORK  
; APPLICANT: HENRY, CHEN  
; TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A  
; TITLE OF INVENTION: SIGNALLING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE  
; TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/476,509B  
; FILING DATE: 01-DEC-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-101 CIP  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 448 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-476-509B-2

Query Match 87.5%; Score 21; DB 3; Length 448;  
Best Local Similarity 66.7%; Pred. No. 1.1e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|||  
Db 139 GTLTPS 144

RESULT 39  
US-08-348-518C-2  
; Sequence 2, Application US/08348518C  
; Patent No. 6022740  
; GENERAL INFORMATION:  
; APPLICANT: SUDOL, MARIUS  
; APPLICANT: PEER, BORK  
; APPLICANT: HENRY, CHEN  
; TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A

```
; TITLE OF INVENTION:  SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE
; TITLE OF INVENTION:  PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES
; TITLE OF INVENTION:  THEREOF
; NUMBER OF SEQUENCES:  26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Klauber & Jackson
; STREET:  411 Hackensack Avenue
; CITY:  Hackensack
; STATE:  New Jersey
; COUNTRY:  USA
; ZIP:  07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/348,518C
; FILING DATE:  01-DEC-1994
; CLASSIFICATION:  514
; ATTORNEY/AGENT INFORMATION:
; NAME:  Jackson Esq., David A.
; REGISTRATION NUMBER:  26,742
; REFERENCE/DOCKET NUMBER:  600-1-101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  201 487-5800
; TELEFAX:  201 343-1684
; TELEX:  133521
; INFORMATION FOR SEQ ID NO:  2:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  486 amino acids
; TYPE:  amino acid
; TOPOLOGY:  linear
; MOLECULE TYPE:  protein
; US-08-348-518C-2

Query Match      87.5%; Score 21; DB 3; Length 486;
Best Local Similarity 66.7%; Pred. No. 1.2e+03;
Matches  4; Conservative  0; Mismatches  2; Indels  0; Gaps  0;

QY      1 gtxxps 6
      || ||
Db      139 GTLTPS 144

RESULT 40
US-08-794-795-2
; Sequence 2, Application US/08794795
; Patent No. 5916766
; GENERAL INFORMATION:
; APPLICANT:  Elshourlagy, Nabil
; APPLICANT:  Adamou, John
; APPLICANT:  Gross, Mitchell
; APPLICANT:  Lysko, Paul
; TITLE OF INVENTION:  Human Macro Scavenger Rec
; TITLE OF INVENTION:  eptor
; NUMBER OF SEQUENCES:  9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  SmithKline Beecham Corporation
; STREET:  709 Swedeland Road
; CITY:  King of Prussia
; STATE:  PA
; COUNTRY:  USA
; ZIP:  19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Diskette
; COMPUTER:  IBM Compatible
; OPERATING SYSTEM:  DOS
; SOFTWARE:  FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/794,795
; FILING DATE:  04-FEB-1997

Query Match      87.5%; Score 21; DB 2; Length 495;
Best Local Similarity 66.7%; Pred. No. 1.2e+03;
Matches  4; Conservative  0; Mismatches  2; Indels  0; Gaps  0;

QY      1 gtxxps 6
      || ||
Db      162 GTTGPS 167

RESULT 41
US-09-249-200-2
; Sequence 2, Application US/09249200
; Patent No. 6197931
; GENERAL INFORMATION:
; APPLICANT:  ELSHOURBAGY, NABIL
; APPLICANT:  ADAMOU, JOHN
; APPLICANT:  GROSS, MITCHELL
; APPLICANT:  LYSKO, PAUL
; TITLE OF INVENTION:  HUMAN MARCO SCAVENGER RECEPTOR
; NUMBER OF SEQUENCES:  9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  Ratner & Prestia
; STREET:  P.O. Box 980
; CITY:  Valley Forge
; STATE:  PA
; COUNTRY:  USA
; ZIP:  19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Diskette
; COMPUTER:  IBM Compatible
; OPERATING SYSTEM:  DOS
; SOFTWARE:  FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/09/249,200
; FILING DATE:  12-FEB-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  08/794,795
; FILING DATE:  04-FEB-1997
; APPLICATION NUMBER:  60/017,699
; FILING DATE:  23-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME:  Prestia, Paul F
; REGISTRATION NUMBER:  23,031
; REFERENCE/DOCKET NUMBER:  ATG-50009-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  610-407-0700
; TELEFAX:  610-407-0700
; TELEX:  846169
; INFORMATION FOR SEQ ID NO:  2:
; SEQUENCE CHARACTERISTICS:
```

; LENGTH: 495 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-249-200-2

Query Match 87.5%; Score 21; DB 4; Length 495;  
Best Local Similarity 66.7%; Pred. No. 1.2e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 162 GTTGPS 167

## RESULT 42

US-09-031-392-4  
; Sequence 4, Application US/09031392  
; Patent No. 5942398  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Weng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES  
; TITLE OF INVENTION: ENCODING GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 534 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-031-392-4

Query Match 87.5%; Score 21; DB 2; Length 534;  
Best Local Similarity 66.7%; Pred. No. 1.3e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 87 GTLAPS 92

## RESULT 43

US-09-299-549-4  
; Sequence 4, Application US/09299549  
; Patent No. 6136547  
; GENERAL INFORMATION:

; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Weng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; TITLE OF INVENTION: GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/299,549  
; FILING DATE: 26-APR-1999  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/031,392  
; FILING DATE: 26-FEB-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meiklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 534 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-299-549-4

Query Match 87.5%; Score 21; DB 4; Length 534;  
Best Local Similarity 66.7%; Pred. No. 1.3e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 87 GTLAPS 92

## RESULT 44

US-09-610-417-4  
; Sequence 4, Application US/09610417  
; Patent No. 6346374  
; GENERAL INFORMATION:  
; APPLICANT: Tartaglia, Louis A.  
; APPLICANT: Weng, Xun  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING  
; GLUTEX AND USES THEREOF  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/610,417

; FILING DATE: 05-Jul-2000  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/299,549  
; FILING DATE: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meklejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/072002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 534 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-610-417-4

Query Match 87.5%; Score 21; DB 4; Length 534;  
Best Local Similarity 66.7%; Pred. No. 1.3e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 87 GTLAPS 92

RESULT 45  
US-09-413-814-4  
; Sequence 4; Application US/09413814  
; Patent No. 6225064  
; GENERAL INFORMATION:  
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
; APPLICANT: Bristol-Myers Squibb, Co.  
; APPLICANT: Beyer, Stefan  
; APPLICANT: Bloeker, Helmut  
; APPLICANT: Brandt, Petra  
; APPLICANT: Cino, Paul M  
; APPLICANT: Dougherty, Brian A  
; APPLICANT: Goldberg, Steven L  
; APPLICANT: Hofle, Gerhard  
; APPLICANT: Mueller, Joachim  
; APPLICANT: Reichenbach, Hans  
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or  
; FILE REFERENCE: PCT/US 99/23535  
; CURRENT APPLICATION NUMBER: US/09/413,814  
; CURRENT FILING DATE: 1999-10-07  
; EARLIER APPLICATION NUMBER: DE 198 46 493.2  
; EARLIER FILING DATE: 1998-10-09  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 577  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-413-814-4

Query Match 87.5%; Score 21; DB 4; Length 577;  
Best Local Similarity 66.7%; Pred. No. 1.4e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 371 GTALPS 376

RESULT 46

US-08-843-530B-28  
; Sequence 28; Application US/08843530B  
; Patent No. 5939306  
; GENERAL INFORMATION:  
; APPLICANT: Selitrennikoff, Claude  
; APPLICANT: Agnan, Jacqueline  
; APPLICANT: Alex, Lisa A.  
; APPLICANT: Simon, Melvin I.  
; TITLE OF INVENTION: Osmosensing Histidine Kinases  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medien & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/843,530B  
; FILING DATE: 16-APR-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MacKnight, Kamrin T.  
; REGISTRATION NUMBER: 38,230  
; REFERENCE/DOCKET NUMBER: UTC-02717  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 28:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 732 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
US-08-843-530B-28

Query Match 87.5%; Score 21; DB 2; Length 732;  
Best Local Similarity 66.7%; Pred. No. 1.7e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 81 GTLSPS 86

RESULT 47  
US-08-836-943-2  
; Sequence 2; Application US/08836943  
; Patent No. 5965391  
; GENERAL INFORMATION:  
; APPLICANT: Reinscheid, Dieter  
; APPLICANT: Eikmanns, Bernhard  
; APPLICANT: Sahm, Hermann  
; TITLE OF INVENTION: DNA WHICH REGULATES GENE EXPRESSION IN  
; TITLE OF INVENTION: CORYNEFORM BACTERIA  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Firm of Karl F. Ross, PC  
; STREET: 5676 Riverdale Ave.  
; CITY: Bronx  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10471  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/836,943  
; FILING DATE: 08-MAY-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Myers, Jonathan  
; REGISTRATION NUMBER: 26,963  
; REFERENCE/DOCKET NUMBER: 20357  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (718) 884-6600  
; TELEFAX: 718/601-1099  
; TELEX: 620428  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 739 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-836-943-2

Query Match 87.5%; Score 21; DB 2; Length 739;  
Best Local Similarity 66.7%; Pred. No. 1.7e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 716 GTKSPS 721

RESULT 48  
US-08-231-193A-42  
; Sequence 42, Application US/08231193A  
; Patent No. 5849895  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR  
; TITLE OF INVENTION: SUBUNITS, NUCLEIC ACIDS ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/231,193A  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/052,459  
; FILING DATE: 20-APR-1993  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 774 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-231-193A-42

Query Match 87.5%; Score 21; DB 2; Length 774;  
Best Local Similarity 66.7%; Pred. No. 1.8e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 292 GTVSPS 297

RESULT 49  
US-08-486-273A-42  
; Sequence 42, Application US/08486273A  
; Patent No. 5985586  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA  
; TITLE OF INVENTION: ENCODING SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486,273A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/231,193  
; FILING DATE: 20-APR-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9383B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 774 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-486-273A-42

Query Match 87.5%; Score 21; DB 2; Length 774;  
Best Local Similarity 66.7%; Pred. No. 1.8e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 292 GTVSPS 297

RESULT 50  
US-08-480-474-42  
; Sequence 42, Application US/08480474  
; Patent No. 6033865  
; GENERAL INFORMATION:  
; APPLICANT: Daggett, Lorrie P.  
; APPLICANT: Ellis, Steven B.  
; APPLICANT: Liaw, Chen W.  
; APPLICANT: Lu, Chin-Chun  
; TITLE OF INVENTION: HUMAN N-METHYL-D-ASPARTATE RECEPTOR SUBUNITS, DNA ENCODING  
; TITLE OF INVENTION: SAME AND USES THEREFOR  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1660 Union Street  
; CITY: San Diego  
; STATE: CA  
; COUNTRY: U.S.A.  
; ZIP: 92101-2926  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/480,474  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 6362-9382B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619-238-0999  
; TELEFAX: 619-238-0062  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 774 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-480-474-42

Query Match 87.5%; Score 21; DB 3; Length 774;  
Best Local Similarity 66.7%; Pred. No. 1.8e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|||  
Db 292 GTVSPS 297

Search completed: September 3, 2002, 09:25:31  
Job time: 23 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 09:25:08 ; Search time 14.73 Seconds  
(without alignments)  
39.140 Million cell updates/sec

Title: BASK-853-CLAIM5  
Perfect score: 24  
Sequence: 1 gtxxps 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 50 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	91.7	89	2 E83141	hypothetical prote
2	22	91.7	101	2 S30493	Spl protein - mous
3	22	91.7	105	2 G72515	hypothetical prote
4	22	91.7	111	2 A85695	hypothetical prote
5	22	91.7	135	2 JC2424	hypothetical 14.3K
6	22	91.7	194	2 H90836	hypothetical prote
7	22	91.7	196	2 I39698	blue copper-bindin
8	22	91.7	196	2 T51838	blue copper bindin
9	22	91.7	202	2 B82133	probable outer mem
10	22	91.7	254	2 A83345	probable thioester
11	22	91.7	255	2 E42725	nitrite hydratase
12	22	91.7	275	2 S47325	myoD protein - zeb
13	22	91.7	280	2 H71320	hypothetical prote
14	22	91.7	285	2 H85356	hypothetical prote
15	22	91.7	289	2 T17957	hypothetical prote
16	22	91.7	316	2 T34553	hypothetical prote
17	22	91.7	319	2 S62196	hypothetical prote
18	22	91.7	379	2 JN0013	synaptic vesicle m
19	22	91.7	381	2 S12223	naringenin-chalcon
20	22	91.7	383	2 E87549	type IV secretion
21	22	91.7	389	1 SIPJCU	naringenin-chalcon
22	22	91.7	389	2 JC5136	naringenin-chalcon
23	22	91.7	398	2 S42523	naringenin-chalcon
24	22	91.7	410	2 S12224	naringenin-chalcon
25	22	91.7	418	2 F75587	probable glycosylt
26	22	91.7	419	1 SVPJCD	naringenin-chalcon
27	22	91.7	419	2 T36272	hypothetical prote
28	22	91.7	421	2 T41156	probable WD repeat
29	22	91.7	435	2 T24477	hypothetical prote

30	22	91.7	460	2 S34969	outer membrane por
31	22	91.7	460	2 T23087	hypothetical prote
32	22	91.7	507	2 T26809	hypothetical prote
33	22	91.7	518	2 F75460	hypothetical prote
34	22	91.7	522	2 T40520	hypothetical prote
35	22	91.7	542	2 T48488	hypothetical prote
36	22	91.7	550	2 B70583	probable transposa
37	22	91.7	574	2 T29005	hypothetical prote
38	22	91.7	593	2 A96783	unknown protein F2
39	22	91.7	597	2 T35746	hypothetical prote
40	22	91.7	629	2 S20516	dnaK-type molecula
41	22	91.7	651	2 A96781	unknown protein F9
42	22	91.7	691	2 B75622	hypothetical prote
43	22	91.7	693	2 T15152	hypothetical prote
44	22	91.7	696	2 A29635	transcription fact
45	22	91.7	756	2 D96527	protein F27J15.24
46	22	91.7	788	2 JS0747	regulatory protein
47	22	91.7	848	2 T28055	hypothetical prote
48	22	91.7	875	2 T20429	hypothetical prote
49	22	91.7	886	1 A47521	capsid protein - g
50	22	91.7	963	2 T40290	hypothetical prote

ALIGNMENTS

RESULT 1

E83141  
hypothetical protein PA4033 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: E83141  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: E83141  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-89 <STO>  
A:Cross-references: GB:AE004820; GB:AE004091; NID:g9950223; PIDN:AAG07420.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4033

Query Match 91.7%; Score 22; DB 2; Length 89;  
Best Local Similarity 66.7%; Pred. No. 72;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 gtxxps 6  
|| ||  
Db 48 GTASPS 53

RESULT 2

S30493  
Spl protein - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C>Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 05-Nov-1999  
C:Accession: S30493  
R:Chestier, A.; Charnay, P.  
A:Title: Difference in the genomic organizations of the related transcription factors  
A:Reference number: S30493; MUID:92338398  
A:Accession: S30493  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-101 <CHE>  
A:Cross-references: EMBL:X60136; NID:g54158; PIDN:CAA42721.1; PID:e38120; PID:g133426



F:79-113/Disulfide bonds: #status predicted

Query Match 91.7%; Score 22; DB 2; Length 196;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 158 GTTTPS 163

RESULT 8

T51838  
blue copper binding protein homolog [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 17-Nov-2000

C:Accession: T51838

R:Yang, K.Y.; Kim, C.S.; Cho, B.H.

submitted to the EMBL Data Library, October 1998

A:Description: Characterization of a wound-inducible Arabidopsis gene encoding a protein

A:Reference number: Z25481

A:Accession: T51838

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-196 <YAN>

A:Cross-references: EMBL:Y18227; PIDN:CAA77089.1

C:Genetics:

A:Gene: AWI 32

C:Superfamily: plastocyanin

C:Keywords: copper binding

Query Match 91.7%; Score 22; DB 2; Length 196;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 158 GTTTPS 163

RESULT 9

B82133  
probable outer membrane lipoprotein Slp VC1987 [imported] - Vibrio cholerae (strain N169

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 23-Mar-2001

C:Accession: B82133

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F.

1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: B82133

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-202 <HEI>

A:Cross-references: GB:AE004273; GB:AE003852; NID:9656517; PIDN:AAF95135.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC1987

A:Map position: 1

C:Superfamily: rnd protein

Query Match 91.7%; Score 22; DB 2; Length 202;  
Best Local Similarity 66.7%; Pred. No. 1.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 134 GTTAPS 139

RESULT 10

A83345

Probable thioesterase PA2411 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: A83345

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337

A:Accession: A83345

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <STO>

A:Cross-references: GB:AE004668; GB:AE004091; NID:g9948446; PIDN:AAG05799.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2411

C:Superfamily: oleoyl-[acyl-carrier-protein] hydrolase; oleoyl-[acyl-carrier-protein]

Query Match 91.7%; Score 22; DB 2; Length 254;

Best Local Similarity 66.7%; Pred. No. 1.9e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 108 GTAAAPS 113

RESULT 11

E42725

nitrile hydratase 3'-hypothetical protein orfE - Pseudomonas chlororaphis (strain B23

C:Species: Pseudomonas chlororaphis

C:Date: 10-Mar-1994 #sequence\_revision 10-Mar-1994 #text\_change 08-Oct-1999

C:Accession: E42725

R:Nishiyama, M.; Horinouchi, S.; Kobayashi, M.; Nagasawa, T.; Yamada, H.; Beppu, T.

J. Bacteriol. 173, 2465-2472, 1991

A:Title: Cloning and characterization of genes responsible for metabolism of nitrile

A:Reference number: A42725; MUID:91193202

A:Accession: E42725

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-255 <NIS>

A:Cross-references: GB:D90216; NID:g216850; PIDN:BAA14248.1; PID:d1014953; PID:g21602

Query Match 91.7%; Score 22; DB 2; Length 255;

Best Local Similarity 66.7%; Pred. No. 1.9e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 103 GTAAAPS 108

RESULT 12

S47325

myoD protein - zebra fish

C:Species: Brachydanio rerio (zebra fish)

C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 08-Dec-2000

C:Accession: S47325

R:Weinberg, E.S.

submitted to the EMBL Data Library, September 1994

A:Description: Developmental regulation of zebrafish myo D in wild type, no tail, and

A:Reference number: S47325

A:Accession: S47325

A:Status: preliminary

A:Molecule type: mRNA

A;Residues: 1-275 <WEI>  
A;Cross-references: EMBL:Z36945; NID:g535814; PID:g535815  
C;Superfamily: human myogenin

Query Match 91.7%; Score 22; DB 2; Length 275;  
Best Local Similarity 66.7%; Pred. No. 2.e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 253 GTTAPS 258

## RESULT 13

H71320  
hypothetical protein TP0462 - syphilis spirochete  
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
C;Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 05-Nov-1999  
C;Accession: H71320  
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, L.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDevitt, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
Science 281, 375-388, 1998  
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
A;Reference number: A71250; MUID:98332770  
A;Accession: H71320  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-280 <COL>  
A;Cross-references: GB:AF001223; GB:AE000520; NID:g3322745; PID:g332275  
A;Experimental source: strain Nichols  
C;Genetics:  
A;Gene: TP0462

Query Match 91.7%; Score 22; DB 2; Length 280;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 116 GTTTPS 121

## RESULT 14

H85356  
hypothetical protein AT4g30510 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C;Accession: H85356  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory, 1999  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488  
A;Accession: H85356  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-285 <STO>  
A;Cross-references: GB:NC\_001268; NID:g7269952; PIDN:CAB79769.1; GSPDB:GN00140  
C;Genetics:  
A;Gene: AT4g30510  
A;Map position: 4

Query Match 91.7%; Score 22; DB 2; Length 285;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 233 GTSSPS 238

## RESULT 15

TI7957  
hypothetical protein A454L - Chlorella virus PBCV-1  
C;Species: Chlorella virus PBCV-1  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C;Accession: TI7957  
R;Graves, M.V.; Van Etten, J.L.  
submitted to the EMBL Data Library, May 1999  
A;Reference number: Z18806  
A;Accession: TI7957  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-289 <GRA>  
A;Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96822.1  
A;Experimental source: specific host Chlorella strain NC64  
C;Genetics:  
A;Note: A454L  
C;Superfamily: Chlorella virus PBCV-1 hypothetical protein A454L

Query Match 91.7%; Score 22; DB 2; Length 289;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 271 GTATPS 276

## RESULT 16

T34553  
hypothetical protein DKFZp434L1435.1 - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 03-Nov-2000  
C;Accession: T34553  
R;Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, October 1999  
A;Reference number: Z21540  
A;Accession: T34553  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-316 <POU>  
A;Cross-references: EMBL:AL122037  
A;Experimental source: adult testis; clone DKFZp434L1435  
C;Genetics:  
A;Note: DKFZp434L1435.1  
C;Superfamily: valine--tRNA ligase

Query Match 91.7%; Score 22; DB 2; Length 316;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 217 GTAAPS 222

## RESULT 17

S62196  
hypothetical protein 2 - Methanosarcina barkeri  
C;Species: Methanosarcina barkeri  
C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 22-Oct-1999  
C;Accession: S62196  
R;Vorholt, J.A.; Vaupel, M.; Thauer, R.K.  
Eur. J. Biochem. 236, 309-317, 1996  
A;Title: A polyferredoxin with eight [4Fe-4S] clusters as a subunit of molybdenum for  
A;Reference number: S62194; MUID:96184912  
A;Accession: S62196  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-319 <VOR>

A:Cross-references: EMBL:X93084; NID:g1124956; PIDN:CAA63627.1; PID:e212293; PID:g112495  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1995

Query Match 91.7%; Score 22; DB 2; Length 319;  
 Best Local Similarity 66.7%; Pred. No. 2.3e+02; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 66 GTSSPS 71

## RESULT 18

JN0013

Synaptic vesicle membrane protein VAT-1 - Pacific electric ray

C:Species: Torpedo californica (Pacific electric ray)

C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 31-Mar-2000

C:Accession: JN0013

R:Linial, M.; Miller, K.; Scheller, R.H.

Neuron 2, 1265-1273, 1989

A:Title: VAT-1: an abundant membrane protein from torpedo cholinergic synaptic vesicles.

A:Reference number: JN0013; MUID:90166593

A:Accession: JN0013

A:Molecule type: mRNA

A:Residues: 1-379 <LIN>

C:Comment: Synaptic vesicles are responsible for regulating the storage and release of n

C:Superfamily: alcohol dehydrogenase; long-chain alcohol dehydrogenase homology

C:Keywords: glycoprotein; membrane protein; phosphoprotein

F:52-350/Domain: long-chain alcohol dehydrogenase homology <LADH>

F:67,127,147/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:273/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 91.7%; Score 22; DB 2; Length 379;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+02; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 194 GRASPS 199

## RESULT 19

S12223

naringenin-chalcone synthase (EC 2.3.1.74) 1 - tomato

C:Species: Lycopersicon esculentum (tomato)

C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 05-May-2000

C:Accession: S12223

R:O'Neill, S.D.; Tong, Y.; Spoerlein, B.; Forkmann, G.; Yoder, J.I.

Mol. Gen. Genet. 224, 279-288, 1990

A:Title: Molecular genetic analysis of chalcone synthase in Lycopersicon esculentum and

A:Reference number: S12223; MUID:91117196

A:Accession: S12223

A>Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-381 <ONE>

C:Genetics:

A:Map position: 9

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 91.7%; Score 22; DB 2; Length 381;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+02; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 23 GTSTPS 28

## RESULT 20

JC5136

naringenin-chalcone synthase (EC 2.3.1.74) 2 - potato

N:Alternate names: chalcone synthase; CHS

C:Species: Solanum tuberosum (potato)

## E87549

type IV secretion system protein B10, probable [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 10-May-2001

C:Accession: E87549

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: E87549

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-383 <STO>

A:Cross-references: GB:AE005673; NID:g13423963; PIDN:AAK24393.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC2422

C:Superfamily: tumor-inducing plasmid pTic58 virB10 protein

Query Match 91.7%; Score 22; DB 2; Length 383;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+02; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 74 GRASPS 79

## RESULT 21

SYPJCJ

naringenin-chalcone synthase (EC 2.3.1.74) J - garden petunia

N:Alternate names: chalcone synthase

C:Species: Petunia x hybrida (garden petunia)

C>Date: 30-Sep-1991 #sequence\_revision 17-Mar-2000 #text\_change 05-May-2000

C:Accession: D72821; JS0309

R:Koes, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.

Gene 81, 245-257, 1989

A:Title: Cloning and molecular characterization of the chalcone synthase multigene fa

A:Reference number: JS0308; MUID:90034197

A:Accession: D72821

A:Molecule type: DNA

A:Residues: 1-389 <KOE>

A:Cross-references: EMBL:X14597; NID:g20535; PIDN:CAA32737.1; PID:g20536

A:Experimental source: strain Violet 30, leaf

A:Accession: JS0309

A:Molecule type: DNA

A:Residues: 1-50,'D',52-74,'V',76-228,'I',230-297,'L',299-389 <KO2>

A:Note: the sequence is revised in GenBank entry PHCHSJ release 114, (PIDN:CAA32737.1

C:Comment: This enzyme plays a central role in the biosynthesis of all classes of fla

C:Genetics:

A:Gene: chsJ

A:Map position: V

A:Note: chsJ is expressed in various floral tissues and UV illuminated seedlings

C:Superfamily: chalcone synthase

C:Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 91.7%; Score 22; DB 1; Length 389;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+02; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 23 GRATPS 28

C;Date: 31-Jan-1997 #sequence\_revision 27-Feb-1997 #text\_change 05-May-2000

C;Accession: JC5136; PC4239

R;Jeon, J.H.; Kim, H.S.; Choi, K.H.; Joung, Y.H.; Joung, H.; Byun, S.M.

Biosci. Biotechnol. Biochem. 60, 1907-1910, 1996

A;Title: Cloning and characterization of one member of the chalcone synthase gene family

A;Reference number: JC5136; MUID:97141614

A;Accession: JC5136

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-389 <JEO1>

A;Cross-references: GB:U47738; NID:g1470059; PIDN:AB05239.1; PID:g1470060

A;Accession: PC4239

A;Status: preliminary

A;Molecule type: protein

A;Residues: 158-165;367-373 <JEO2>

C;Comment: This enzyme is important in the biosynthesis of all classes of flavonoids in

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A

Query Match 91.7%; Score 22; DB 2; Length 389;

Best Local Similarity 66.7%; Pred. No. 2.8e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 23 GTATPS 28

RESULT 23

S42523

naringenin-chalcone synthase (EC 2.3.1.74) - parsley

N;Alternate names: chalcone synthase

C;Species: Petroselinum crispum (parsley)

C;Date: 07-Sep-1994 #sequence\_revision 10-Nov-1995 #text\_change 05-May-2000

C;Accession: S42523

R;Reimold, U.; Kroeger, M.; Kreuzaler, F.; Hahlbrock, K.

EMBO J. 2, 1801-1805, 1983

A;Title: Coding and 3' non-coding nucleotide sequence of chalcone synthase mRNA and ass

A;Reference number: S42523

A;Accession: S42523

A;Molecule type: mRNA

A;Residues: 1-398 <REI>

A;Cross-references: EMBL:V01538; NID:g20513; PIDN:CAA24779.1; PID:g20514

A;Note: the source is designated as Petroselinum hortense

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 91.7%; Score 22; DB 2; Length 398;

Best Local Similarity 66.7%; Pred. No. 2.9e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 28 GTATPS 33

RESULT 24

S12224

naringenin-chalcone synthase (EC 2.3.1.74) 2 - tomato

C;Species: Lycopersicon esculentum (tomato)

C;Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 05-May-2000

C;Accession: S12224

R;O'Neill, S.D.; Tong, Y.; Spoerlein, B.; Forkmann, G.; Yoder, J.I.

Mol. Gen. Genet. 224, 279-288, 1990

A;Title: Molecular genetic analysis of chalcone synthase in Lycopersicon esculentum and

A;Reference number: S12223; MUID:91117196

A;Accession: S12224

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-410 <ONE>

C;Genetics:

A;Map position: 5

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 91.7%; Score 22; DB 2; Length 410;

Best Local Similarity 66.7%; Pred. No. 3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 23 GTATPS 28

RESULT 25

F75587

probable glycosyltransferase - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C;Accession: F75587

R;White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896

A;Accession: F75587

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-418 <WHI>

A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12451.1; PID:g646

A;Experimental source: strain R1

C;Genetics:

A;Gene: DRA0329

A;Map position: 2

C;Superfamily: glycosyltransferase

Query Match 91.7%; Score 22; DB 2; Length 418;

Best Local Similarity 66.7%; Pred. No. 3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 365 GTATPS 370

RESULT 26

SYPUCD

naringenin-chalcone synthase (EC 2.3.1.74) D - garden petunia

N;Alternate names: chalcone synthase

C;Species: Petunia x hybrida (garden petunia)

C;Date: 30-Sep-1991 #sequence\_revision 17-Mar-2000 #text\_change 21-Jul-2000

C;Accession: A72821; JS0312

R;Koes, R.E.; Spelt, C.E.; van den Elzen, P.J.M.; Mol, J.N.M.

Gene 81, 245-257, 1989

A;Title: Cloning and molecular characterization of the chalcone synthase multigene fa

A;Reference number: JS0308; MUID:90034197

A;Accession: A72821

A;Molecule type: DNA

A;Residues: 1-419 <KOE>

A;Cross-references: GB:X14593; NID:g20528; PIDN:CAA32733.1; PID:g20529

A;Experimental source: strain Violet 30, leaf

A;Accession: JS0312

A;Molecule type: DNA

A;Residues: 1-117, 'D', '119', 'H', '121-313', 'DI', '316-419 <KO2>

A;Note: the sequence was revised in GenBank entry PHCHSD, release 114, (PIDN:CAA32733

C;Comment: This enzyme plays a central role in the biosynthesis of all classes of fla

C;Genetics:

A;Gene: chsd

A;Map position: V

C;Superfamily: chalcone synthase

C;Keywords: acyltransferase; coenzyme A; flavonoid biosynthesis

Query Match 91.7%; Score 22; DB 1; Length 419;  
 Best Local Similarity 66.7%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 23 GTATPS 28

RESULT 27  
 T36272  
 hypothetical protein SCE68.22 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T36272  
 R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, June 1999  
 A:Reference number: Z21576  
 A:Accession: T36272  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-419 <MUR>  
 A:Cross-references: EMBL:AL079345; PIDN:CAB45358.1; GSPDB:GN00070; SCODEB:SCE68.22  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCODEB:SCE68.22

Query Match 91.7%; Score 22; DB 2; Length 419;  
 Best Local Similarity 66.7%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 380 GTASPS 385

RESULT 28  
 T41156  
 probable WD repeat transcription regulation protein - fission yeast (Schizosaccharomyces  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 07-Dec-1999  
 C:Accession: T41156  
 R:Rilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, October 1998  
 A:Reference number: Z21973  
 A:Accession: T41156  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-421 <HIL>  
 A:Cross-references: EMBL:AL031907; PIDN:CRA21427.1; GSPDB:GN00068; SPDB:SPCC18.13  
 A:Experimental source: strain 972h-; cosmid c18  
 C:Genetics:  
 A:Gene: SPDB:SPCC18.13  
 A:Map position: 3  
 A:Introns: 227/2

Query Match 91.7%; Score 22; DB 2; Length 421;  
 Best Local Similarity 66.7%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 410 GTSSPS 415

RESULT 29  
 T24477  
 hypothetical protein T04H1.7 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
 C:Accession: T24477  
 R:Harris, B.  
 submitted to the EMBL Data Library, August 1996  
 A:Reference number: Z19896  
 A:Accession: T24477  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-435 <WIL>  
 A:Cross-references: EMBL:Z78200; PIDN:CAB01584.1; GSPDB:GN00023; CESP:T04H1.7  
 A:Experimental source: clone T04H1  
 C:Genetics:  
 A:Gene: CESP:T04H1.7  
 A:Map position: 5  
 A:Introns: 54/3; 82/3; 193/1; 239/3; 282/3; 360/1  
 C:Superfamily: glucuronosyltransferase

Query Match 91.7%; Score 22; DB 2; Length 435;  
 Best Local Similarity 66.7%; Pred. No. 3.1e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 300 GTTTPS 305

RESULT 30  
 S34969  
 outer membrane porin OprE precursor PA0291 [imported] - Pseudomonas aeruginosa (strai  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 15-Jun-2001  
 C:Accession: S34969; E83608  
 R:Yamano, Y.; Nishikawa, T.; Komatsu, Y.  
 Mol. Microbiol. 8, 993-1004, 1993  
 A:Title: Cloning and nucleotide sequence of anaerobically induced porin protein El (O  
 A:Reference number: S34969; MUID:93360827  
 A:Accession: S34969  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-460 <YAM>  
 A:Cross-references: GB:D12711; NID:g433417; PIDN:BAA02207.1; PID:g433418  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;  
 Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
 A:Reference number: A82950; MUID:20437337  
 A:Accession: E83608  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-460 <STO>  
 A:Cross-references: GB:AE004467; GB:AE004091; NID:g9946133; PIDN:AAG03680.1; GSPDB:GN  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: OprE; PA0291  
 C:Superfamily: Pseudomonas aeruginosa outer membrane porin PA0958

Query Match 91.7%; Score 22; DB 2; Length 460;  
 Best Local Similarity 66.7%; Pred. No. 3.3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 Db 54 GTASPS 59

RESULT 31  
 T23087  
 hypothetical protein H13N06.2 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T23087

R:Renard, N.

Submitted to the EMBL Data Library, October 1997

A:Reference number: Z19673

A:Accession: T23087

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-460 <WIL>

A:Cross-references: EMBL:Z99942; PIDN:CAB17068.1; GSPDB:GN00028; CESP:H13N06.2

A:Experimental source: clone H13N06

C:Genetics:

A:Gene: CESP:H13N06.2

A:Map position: X

A:Introns: 17/1; 294/1; 331/3; 364/1; 414/2; 432/2

Query Match

Best Local Similarity 91.7%; Score 22; DB 2; Length 460;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

Db 86 GTSTPS 91

RESULT 32

T26809.

hypothetical protein Y66A7A.8 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T26809; T27293

R:Steward, C.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20269

A:Accession: T26809

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-507 <WIL>

A:Cross-references: EMBL:AL032627; PIDN:CAA21552.1; CESP:Y66A7A.8

A:Experimental source: clone Y41C4A

R:Matthews, L.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20338

A:Accession: T27293

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-507 <WT2>

A:Cross-references: EMBL:AL032622; PIDN:CAA21503.1; GSPDB:GN00021; CESP:Y66A7A.8

A:Experimental source: clone Y66A7A

C:Genetics:

A:Gene: CESP:Y66A7A.8

A:Map position: 3

A:Introns: 29/3; 93/1; 145/2; 314/3; 358/3

Query Match

Best Local Similarity 91.7%; Score 22; DB 2; Length 507;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

Db 425 GTSSPS 430

RESULT 33

F75460

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: F75460

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: F75460

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-518 <WHI>

A:Cross-references: GB:AE001943; GB:AE000513; NID:G6458624; PIDN:AAF10486.1; PID:G645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR0903

A:Map position: 1

Query Match

Best Local Similarity 91.7%; Score 22; DB 2; Length 518;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

Db 179 GTTTPS 184

RESULT 34

T40520

hypothetical protein SPBC530.04 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T40520

R:Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.

submitted to the EMBL Data Library, May 1998

A:Reference number: Z21934

A:Accession: T40520

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-522 <LYN>

A:Cross-references: EMBL:AL023634; PIDN:CAA19170.1; GSPDB:GN00067; SPDB:SPBC530.04

A:Experimental source: strain 972h-; cosmid c530

C:Genetics:

A:Gene: SPDB:SPBC530.04

A:Map position: 2

A:Introns: 20/3

Query Match

Best Local Similarity 91.7%; Score 22; DB 2; Length 522;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

Db 41 GTSAPS 46

RESULT 35

T48488

hypothetical protein T28J14.110 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000

C:Accession: T48488

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.;

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24493

A:Accession: T48488

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-542 <BEV>

A:Cross-references: EMBL:AL163652

A:Experimental source: cultivar Columbia; BAC clone T28J14

C:Genetics:

A:Map position: 5

A:Introns: 162/3; 227/1; 251/1; 299/2; 355/3; 377/2; 397/3; 419/3; 435/3; 449/2; 487/

A:Note: T28J14.110

Query Match 91.7%; Score 22; DB 2; Length 542;  
Best Local Similarity 66.7%; Pred. No. 3.8e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 400 GTTSPS 405  
  
RESULT 36  
B70583  
Probable transposase for IS1535 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C:Accession: B70583  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987  
A:Accession: B70583  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-550 <COL>  
A:Cross-references: GB:J95210; GB:AL123456; NID:g3261757; PIDN:CAB08504.1; PID:e315222;  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: Rv0922

Query Match 91.7%; Score 22; DB 2; Length 550;  
Best Local Similarity 66.7%; Pred. No. 3.9e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 497 GTTAPS 502

RESULT 37  
T29005  
Hypothetical protein ZK328.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T29005  
R:Favell, A.  
submitted to the EMBL Data Library, March 1996  
A:Description: The sequence of C. elegans cosmid ZK328.  
A:Reference number: Z20552  
A:Accession: T29005  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-574 <FAV>  
A:Cross-references: EMBL:U50193; PIDN:AAA91246.1; CESP:ZK328.3  
A:Experimental source: strain Bristol N2  
C:Genetics:  
A:Gene: CESP:ZK328.3  
A:Introns: 63/1; 140/2; 419/1; 436/2; 488/2; 516/1; 545/2

Query Match 91.7%; Score 22; DB 2; Length 574;  
Best Local Similarity 66.7%; Pred. No. 4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 284 GTSTPS 289

RESULT 38

A96783  
unknown protein F22H5.2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: A96783  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, ansen, N.F.; Hughes, B.; Huizar, L. Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzlia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719  
A:Accession: A96783  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-593 <STO>  
A:Cross-references: GB:AE005173; NID:gl00923269; PIDN:AAG12682.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F22H5.2  
A:Map position: 1

Query Match 91.7%; Score 22; DB 2; Length 593;  
Best Local Similarity 66.7%; Pred. No. 4.2e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 22 GTSSPS 27

RESULT 39  
T35746  
Hypothetical protein SC7H2.15c - Streptomyces coelicolor  
C:Species: Streptomyces coelicolor  
C:Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C:Accession: T35746  
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, submitted to the EMBL Data Library, August 1999  
A:Reference number: Z21588  
A:Accession: T35746  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-597 <SAU>  
A:Cross-references: EMBL:AL109732; PIDN:CAB52057.1; GSPDB:GN00070; SCOEDB:SC7H2.15c  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOEDB:SC7H2.15c

Query Match 91.7%; Score 22; DB 2; Length 597;  
Best Local Similarity 66.7%; Pred. No. 4.2e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
|| ||  
Db 299 GTATPS 304

RESULT 40  
S20516  
dnaK-type molecular chaperone hsp70, chloroplast - chromophytic alga (Pavlova lutheri N)  
A:Alternate names: heat shock protein 70  
C:Species: Chloroplast Pavlova lutheri  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999  
C:Accession: S20516  
R:Scaramuzzi, C.D.; Stokes, H.W.; Hiller, R.G. Plant Mol. Biol. 18, 467-476, 1992  
A:Title: Heat shock Hsp70 protein is chloroplast-encoded in the chromophytic alga Pav

A:Reference number: S20516; MUID:92163012

A:Accession: S20516

A:Molecule type: DNA

A:Residues: 1-629 <SCA>

A:Cross-references: EMBL:X59555; NID:g20904; PIDN:CAA42154.1; PID:g20905

C:Genetics:

A:Gene: hsp70

A:Genome: chloroplast

C:Function:

A:Description: involved in protein folding and assembling/disassembling of protein complex

C:Superfamily: heat shock protein 70

C:Keywords: ATP; chloroplast; molecular chaperone

Query Match 91.7%; Score 22; DB 2; Length 629;

Best Local Similarity 66.7%; Pred. No. 4.4e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6

|| || ||

Db 33 GTTTPS 38

RESULT 41

A96781

unknown protein F9E10.5 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: A96781

R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

A:Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: A96781

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-651 <STO>

A:Cross-references: GB:AE005173; NID:g6646755; PIDN:AAF21067.1; GSPDB:GN00141

C:Genetics:

A:Gene: F9E10.5

A:Map position: 1

Query Match

Best Local Similarity 91.7%; Score 22; DB 2; Length 651;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6

|| || ||

Db 129 GTSPS 134

RESULT 42

B75622

hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000

C:Accession: B75622

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, R.J.;

M.; Shen, M.; Vanathavan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: B75622

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-691 <WHI>

A:Cross-references: GB:AE001826; NID:g6460827; PIDN:AAF12628.1; PID:g6460924; TIGR:DR

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRB0037

A:Map position: megaplasmid

A:Genome: plasmid

A>Note: plasmid MP1

Query Match

Best Local Similarity 91.7%; Score 22; DB 2; Length 691;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6

|| || ||

Db 563 GTTTPS 568

RESULT 43

T15152

hypothetical protein F35F10.10 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T15152

R:Rohlfing, T.

submitted to the EMBL Data Library, May 1997

A:Description: The sequence of C. elegans cosmid F35F10.

A:Reference number: Z18301

A:Accession: T15152

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-693 <ROH>

A:Cross-references: EMBL:AF002198; NID:g2076898; PID:g2076909; PIDN:AAB53997.1; GSPDB

A:Experimental source: strain Bristol N2; clone F35F10

C:Genetics:

A:Gene: CESP:F35F10.10

A:Map position: 5

A:Introns: 21/2; 65/1; 83/1; 169/3; 193/3; 235/3; 259/3; 289/1; 369/3; 412/3; 495/1;

Query Match

Best Local Similarity 91.7%; Score 22; DB 2; Length 693;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6

|| || ||

Db 486 GTTTPS 491

RESULT 44

A29635

transcription factor Spl - human (fragment)

N:Alternate names: finger protein ZNF76

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 05-Nov-1999

C:Accession: A29635; G44256

R:Kadonaga, J.T.; Carner, K.R.; Maslarsz, F.R.; Tjian, R.

Cell 51, 1079-1090, 1987

A:Title: Isolation of cDNA encoding transcription factor Spl and functional analysis

A:Reference number: A29635; MUID:88080466

A:Accession: A29635

A:Molecule type: mRNA

A:Residues: 1-696 <KAD>

A:Cross-references: GB:J03133; NID:g339517; PIDN:AAA61154.1; PID:g339518

R:Ragoussis, J.; Senger, G.; Mockridg, I.; Sansau, P.; Ruddy, S.; Dudley, K.; Sheer

Genomics 14, 673-679, 1992

A:Title: A testis-expressed Zn finger gene (ZNF76) in human 6p21.3 centromeric to the

A:Reference number: A44256; MUID:93052398

A:Accession: G44256

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A;Residues: 569-598 <RAG>  
A;Experimental source: T-cell line CEM  
A;Note: sequence extracted from NCBI backbone (NCBIP:125980)  
C;Genetics:  
A;Gene: GDB:SP1  
A;Cross-references: GDB:127453; OMIM:189906  
A;Map position: 19q13.1-19q13.3  
C;Keywords: DNA binding; transcription regulation; zinc finger

Query Match 91.7%; Score 22; DB 2; Length 696;  
Best Local Similarity 66.7%; Pred. No. 4.8e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 647 GTATPS 652

RESULT 45  
D96527  
protein F77J15.24 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cross)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: D96527  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719  
A;Accession: D96527  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-756 <STO>  
A;Cross-references: GB:AE005173; NID:g7770335; PIDN:AAF69705.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F77J15.24  
A;Map position: 1

Query Match 91.7%; Score 22; DB 2; Length 756;  
Best Local Similarity 66.7%; Pred. No. 5.2e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 118 GTSSPS 123

RESULT 46  
JS0747  
regulatory protein Sp1 - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 01-Dec-2000  
C;Accession: JS0747; S25287  
R;Imataka, H.; Sogawa, K.; Yasumoto, K.; Kikuchi, Y.; Sasano, K.; Kobayashi, A.; Hayami,  
submitted to JIPID, September 1992  
A;Reference number: JS0747  
A;Accession: JS0747  
A;Molecule type: mRNA  
A;Residues: 1-788 <IMA>  
A;Cross-references: DBJ:D12768; NID:g220911; PIDN:BA02235.1; PID:d1002730; PID:g220912  
R;Imataka, H.; Sogawa, K.; Yasumoto, K.; Kikuchi, Y.; Sasano, K.; Kobayashi, A.; Hayami,  
EMBO J. 11, 3663-3671, 1992  
A;Title: Two regulatory proteins that bind to the basic transcription element (BTE), a G  
A;Reference number: S25287; MUID:93010958  
A;Accession: S25287

A;Molecule type: mRNA  
A;Residues: 1-122, 'L', '124-311, 'A', 313-788 <IM2>  
C;Keywords: DNA binding; transcription regulation

Query Match 91.7%; Score 22; DB 2; Length 788;  
Best Local Similarity 66.7%; Pred. No. 5.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 739 GTATPS 744

RESULT 47  
T28055  
hypothetical protein ZK858.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T28055  
R;White, S.  
submitted to the EMBL Data Library, September 1996  
A;Reference number: Z20462  
A;Accession: T28055  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-848 <WIL>  
A;Cross-references: EMBL:Z79759; PIDN:CAB02138.1; GSPDB:GN00019; CESP:ZK858.1  
C;Experimental source: clone ZK858  
C;Genetics:  
A;Gene: CESP:ZK858.1  
A;Map position: 1  
A;Introns: 26/3; 82/3; 130/3; 195/3; 241/3; 613/1; 762/1

Query Match 91.7%; Score 22; DB 2; Length 848;  
Best Local Similarity 66.7%; Pred. No. 5.8e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 630 GTAAPS 635

RESULT 48  
T20429  
hypothetical protein E03A3.6 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C;Accession: T20429  
R;Gardner, A.  
submitted to the EMBL Data Library, October 1994  
A;Reference number: Z19274  
A;Accession: T20429  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-875 <WIL>  
A;Cross-references: EMBL:Z38112; PIDN:CAA86231.1; GSPDB:GN00021; CESP:E03A3.6  
A;Experimental source: clone E03A3  
C;Genetics:  
A;Gene: CESP:E03A3.6  
A;Map position: 3  
A;Introns: 58/2; 161/3; 200/1; 257/3; 436/3; 629/3; 735/1; 762/3  
C;Superfamily: Caenorhabditis elegans hypothetical protein E03A3.6

Query Match 91.7%; Score 22; DB 2; Length 875;  
Best Local Similarity 66.7%; Pred. No. 5.9e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 827 GTSAPS 832

## RESULT 49

A47521  
capsid protein - giardiavirus GLV  
C:Species: giardiavirus, GLV  
C:Date: 21-Jan-1994 #sequence\_revision 13-Feb-1998 #text\_change 16-Jul-1999  
C:Accession: A47521  
R:Wang, A.L.; Yang, H.M.; Shen, K.A.; Wang, C.C.  
Proc. Natl. Acad. Sci. U.S.A. 90, 8595-8599, 1993  
A:Title: Giardiavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-  
A:Reference number: A47521; MUID:93391401  
A:Contents: host Giardia lamblia  
A:Accession: A47521  
A:Molecule type: genomic RNA  
A:Residues: 1-886 <WAN>  
A:Cross-references: GB:L13218; NID:q1352866; PIDN:RAB01578.1; PID:q1352867  
A:Note: sequence modified after extraction from NCBI backbone (NCBIN:137593, NCBIPI:13759  
C:Superfamily: giardiavirus capsid protein  
C:Keywords: capsid protein

Query Match 91.7%; Score 22; DB 1; Length 886;  
Best Local Similarity 66.7%; Pred. No. 6e+02; 2; Indels 0; Gaps 0;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 796 GTAAPS 801

## RESULT 50

T40290  
hypothetical protein SPBC354.10 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T40290  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.  
submitted to the EMBL Data Library, March 1998  
A:Reference number: Z21918  
A:Accession: T40290  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-963 <WOO>  
A:Cross-references: EMBL:AL022071; PIDN:CAAL7810.1; GSPDB:GN00067; SPDB:SPBC354.10  
A:Experimental source: strain 972h-; cosmid c354  
C:Genetics:  
A:Gene: SPDB:SPBC354.10  
A:Map position: 2

Query Match 91.7%; Score 22; DB 2; Length 963;  
Best Local Similarity 66.7%; Pred. No. 6.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 730 GTAPS 735

Search completed: September 3, 2002, 09:25:53  
Job time: 45 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 09:25:58 ; Search time 10.34 seconds  
(without alignments)  
22.468 Million cell updates/sec

Title: BASK-853-CLAIM5

Perfect score: 24

Sequence: 1 gtxps 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	22	91.7	196	1	BCB_ARATH
2	22	91.7	255	1	YORE_PSECL
3	22	91.7	275	1	MYOD_BRARE
4	22	91.7	312	1	MTR7_MOUSE
5	22	91.7	379	1	VAT1_TORCA
6	22	91.7	389	1	CHS1_LYCES
7	22	91.7	389	1	CHS2_LYCES
8	22	91.7	389	1	CHS2_SOLTU
9	22	91.7	389	1	CHSA_SOLTU
10	22	91.7	389	1	CHSB_SOLTU
11	22	91.7	389	1	CHSJ_PETHY
12	22	91.7	389	1	CHSV_CATRO
13	22	91.7	389	1	CHSV_PRAVU
14	22	91.7	392	1	Y462_TREPA
15	22	91.7	395	1	CHSV_BETVE
16	22	91.7	397	1	CHS2_DAUCA
17	22	91.7	398	1	CHSV_CALCH
18	22	91.7	398	1	CHSV_PETCR
19	22	91.7	419	1	CHSD_PETHY
20	22	91.7	435	1	UGTE_CAEEL
21	22	91.7	629	1	DNAK_PAVLU
22	22	91.7	785	1	SPI_HUMAN
23	22	91.7	788	1	SPI_RAT
24	22	91.7	1050	1	ULK1_HUMAN
25	22	91.7	1262	1	GNRP_MOUSE
26	22	91.7	1377	1	CID_DROME
27	22	91.7	2717	1	ZEP1_HUMAN
28	22	91.7	2769	1	THYG_BOVIN
29	21	87.5	45	1	LYS3_ECOLI
30	21	87.5	45	1	LYS3_SHISO
31	21	87.5	47	1	LYS2_ECOLI
32	21	87.5	47	1	LYS5_ECOLI
33	21	87.5	47	1	LYS6_ECOLI

34	21	87.5	47	1	LYS7_ECOLI
35	21	87.5	47	1	LYS8_ECOLI
36	21	87.5	47	1	LYS9_ECOLI
37	21	87.5	49	1	LYS0_ECOLI
38	21	87.5	52	1	LYS4_ECOLI
39	21	87.5	80	1	PAHO_CHICK
40	21	87.5	94	1	YQGV_BACSU
41	21	87.5	118	1	TRH4_ECOLI
42	21	87.5	122	1	RK14_PORPU
43	21	87.5	170	1	FMF7_ECOLI
44	21	87.5	187	1	OMZ4_ARATH
45	21	87.5	193	1	Y319_MYCPN
46	21	87.5	203	1	COX3_MYCTU
47	21	87.5	215	1	NODE_RHISN
48	21	87.5	225	1	XYN2_ASPNG
49	21	87.5	242	1	PDXJ_NEIMA
50	21	87.5	242	1	PDXJ_NEIMB

#### ALIGNMENTS

RESULT 1

ID	BCB_ARATH	STANDARD;	PRT;	196 AA.
AC	Q07488;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Blue copper protein precursor.			
GN	BCB.			
OS	Arabidopsis thaliana (Mouse-ear cross).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosids II; Brassicales; Brassicaceae; Arabidopsi			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, COLUMBIA, AND CV. K85;			
RX	MEDLINE=94124044; PubMed=8294044;			
RA	van Gysel A., van Montagu M., Inze D.;			
RT	"A negatively light-regulated gene from Arabidopsis thaliana encodes			
RT	a protein showing high similarity to blue copper-binding proteins.";			
RL	Gene 136:79-85(1993).			
CC	-!- FUNCTION: PROBABLY ACTS AS AN ELECTRON CARRIER.			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).			
CC	-!- DEVELOPMENTAL STAGE: MAXIMUM LEVELS ARE FOUND IN 35 DAY OLD			
CC	PLANTLETS WHEN THE ROSETTE IS MATURE, CONSISTING OF 8-10 FULLY			
CC	EXPANDED LEAVES, AND AS THE FLORAL STEM STARTS TO FORM. THIS LEVEL			
CC	REMAINS CONSTANT DURING THE FURTHER LIFE SPAN OF THE PLANT.			
CC	-!- INDUCTION: BY DARK ADAPTATION. THIS GIVES A 20-FOLD INCREASE IN			
CC	EXPRESSION.			
CC	-!- SIMILARITY: CONTAINS 1 PLASTOCYANIN-LIKE DOMAIN.			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; Z15058; CAA78771.1;			
DR	PIR; S25555; S25555.			
DR	HSSP; P29602; IJER.			
DR	InterPro; IPR000923; Copper_blue1.			
DR	InterPro; IPR003245; Cu_bind_like.			
DR	Pfam; PF02298; Cu_bind_like; 1.			
DR	ProDom; PD003122; Cu_bind_like; 1.			
DR	PROSITE; PS00196; COPPER_BLUE; 1.			
KW	Electron transport; Copper; Transmembrane; Signal; Metal-binding;			
KW	Glycoprotein.			
FT	SIGNAL	1	22	POTENTIAL.

```

FT CHAIN          23 196 BLUE COPPER PROTEIN.
FT DOMAIN         23 118 PLASTOCYANIN-LIKE.
FT TRANSMEM      119 143 POTENTIAL.
FT TRANSMEM      173 189 POTENTIAL.
FT DISULFID      79 113 BY SIMILARITY.
FT METAL         66 66 COPPER (BY SIMILARITY).
FT METAL        107 107 COPPER (BY SIMILARITY).
FT METAL        112 112 COPPER (BY SIMILARITY).
FT CARBOHYD      98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE      196 AA; 20071 MW; 432D7270268F047A CRC64;

Query Match          91.7%; Score 22; DB 1; Length 196;
Best Local Similarity 66.7%; Pred. No. 48;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 158 GTTTPS 163

RESULT 2
YORE_PSECL          STANDARD; PRT; 255 AA.
AC Q03003;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein in P47K 3'region (ORFE) (Fragment).
OS Pseudomonas chlororaphis (Pseudomonas aureofaciens).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC Pseudomonas.
OX NCBI_TaxID=333;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B23;
RA MEDLINE=91193202; PubMed=2013568;
RA Nishiyama M., Horinouchi S., Kobayashi M., Nagasawa T., Yamada H.,
RA Beppu T.;
RT "Cloning and characterization of genes responsible for metabolism of
RT nitrite compounds from Pseudomonas chlororaphis B23.";
RL J. Bacteriol. 173:2465-2472(1991).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90216; BAA14248.1; -
DR PIR; E42725; E42725.
DR InterPro; IPR001610; PAC.
DR Pfam; PF00785; PAC; 1.
DR SMART; SM00086; PAC; 1.
KW Hypothetical protein.
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 28905 MW; 073B230C8E38E0D2 CRC64;

Query Match          91.7%; Score 22; DB 1; Length 255;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 103 GTAAPS 108

RESULT 3
MYOD_BRARE          STANDARD; PRT; 275 AA.
ID MYOD_BRARE

```

```

AC Q04077;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Myoblast determination protein 1 homolog (Myogenic factor 1).
GN MYOD.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96152221; PubMed=8565839;
RA Weinberg E.S., Allende M.L., Kelly C.S., Abdelhamid A., Murakami T.,
RA Andermann P., Doerre O.G., Grunwald D.J., Riggleman B.;
RA "Developmental regulation of zebrafish MyoD in wild-type, no tail and
RA spadetail embryos.";
RL Development 122:271-280(1996).
CC -!- FUNCTION: MYOGENESIS.
CC -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC BHLH PROTEIN.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION IS SPECIFIC TO THE MESODERM OF THE
CC GASTRULA.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. "MYOGENIC FACTORS" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z36945; CAA85407.1; -
DR HSPG; P10085; IMDY.
DR ZFIN; ZDB-GENE-980526-561; myod.
DR InterPro; IPR002546; Basic.
DR InterPro; IPR003015; HLH_Myc.
DR InterPro; IPR001092; HLH_dim.
DR Pfam; PF01586; Basic; 1.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00520; BASIC; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HELIX LOOP HELIX; 1.
KW Myogenesis; Differentiation; Developmental protein; Nuclear protein;
KW Transcription regulation; DNA-binding.
FT DOMAIN 84 96 BASIC DOMAIN.
FT DOMAIN 97 136 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
SQ SEQUENCE 275 AA; 30866 MW; 260091DDA756311C CRC64;

Query Match          91.7%; Score 22; DB 1; Length 275;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 253 GTTAPS 258

RESULT 4
MTR7_MOUSE          STANDARD; PRT; 312 AA.
ID MTR7_MOUSE
AC Q922C9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myotubularin related protein 7 (EC 3.1.3.48) (Fragment).
GN MTR7.
OS Mus musculus (Mouse).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98409499; PubMed=9736772;  
 RA Laporte J., Blondeau F., Buj-Bello A., Tentler D., Kretz C., Dahl N.,  
 RA Mandel J.-L.;  
 RT "Characterization of the myotubularin dual specificity phosphatase  
 RT gene family from yeast to human.";  
 RL Hum. Mol. Genet. 7:1703-1712(1998).  
 CC -1- FUNCTION: NOT KNOWN, COULD BE A TYROSINE-PHOSPHATASE.  
 CC -1- SIMILARITY: BELONGS TO THE MYOTUBULARIN-LIKE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF073882; AAC80004.1; -  
 DR MGD; MGI:1891693; Mtmr7.  
 DR InterPro: IPR000340; DS\_phosphatase.  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR SMART: SM00012; PTPC\_DSPC; 1.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 1.  
 KW Hydrolase.  
 FT NON\_TER 1 1  
 FT ACT\_SITE 48 48 BY SIMILARITY.  
 SQ SEQUENCE 312 AA; 36018 MW; 4BB40E0D15BD880C CRC64;  
 Query Match 91.7%; Score 22; DB 1; Length 312;  
 Best Local Similarity 66.7%; Pred. No. 75;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gtxxps 6  
 Db 294 GTASPS 299  
 RESULT 5  
 VAT1\_TORCA  
 ID VAT1\_TORCA STANDARD; PRT; 379 AA.  
 AC P19333;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Synaptic vesicle membrane protein VAT-1.  
 OS Torpedo californica (Pacific electric ray).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Squala; Hypnosquala; Pristiogalea; Batoidae;  
 OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.  
 OX NCBI\_TaxID=7787;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Electric lobe;  
 RX MEDLINE=90166593; PubMed=2483112;  
 RA Linial M., Miller K., Scheller R.H.;  
 RT "VAT-1: an abundant membrane protein from Torpedo cholinergic  
 RT synaptic vesicles.";  
 RL Neuron 2:1265-1273(1989).  
 CC -1- FUNCTION: MAY PLAY A CENTRAL ROLE IN THE FUNCTIONS MEDIATED BY  
 CC SPECIFIC CLASSES OF SYNAPTIC VESICLES.  
 CC -1- SUBCELLULAR LOCATION: Membrane.  
 CC -1- TISSUE SPECIFICITY: CHOLINERGIC SYNAPTIC VESICLES.  
 CC -1- MISCELLANEOUS: SYNTHESIZED IN THE NEURONAL CELL BODIES AND  
 CC TRANSPORTED TO THE TERMINALS. EACH VESICLE CONTAINS APPROXIMATELY  
 CC 28 MOLECULES OF VAT-1.  
 CC -1- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE

CC FAMILY. QUINONE OXIDOREDUCTASE SUBFAMILY. STRONG, TO MAMMALIAN  
 CC VAT-1 HOMOLOGS.  
 CC PIR: JN0013; JN0013.  
 DR InterPro: IPR002085; Adh\_zn\_family.  
 DR InterPro: IPR002364; OOR\_zeta\_crystal.  
 DR Pfam: PF00107; adh\_zinc; 1.  
 DR PROSITE: PS01162; OOR\_ZETA\_CRYSTAL; 1.  
 KW Oxidoreductase; Zinc; Synapse; Membrane; Phosphorylation.  
 FT MOD\_RES 273 273 PHOSPHORYLATION (POTENTIAL).  
 FT SEQUENCE 379 AA; 41593 MW; FA4ADAI7E657F09C CRC64;  
 Query Match 91.7%; Score 22; DB 1; Length 379;  
 Best Local Similarity 66.7%; Pred. No. 90;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gtxxps 6  
 Db 194 GTASPS 199  
 RESULT 6  
 CHSL\_LYCES  
 ID CHSL\_LYCES STANDARD; PRT; 389 AA.  
 AC P23418;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Chalcone synthase 1 (EC 2.3.1.74) (Naringenin-chalcone synthase 1).  
 GN CHSL.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
 OX NCBI\_TaxID=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cotyledon, Hypocotyl, and Leaf;  
 RX MEDLINE=91117196; PubMed=1980524;  
 RA O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;  
 RT "Molecular genetic analysis of chalcone synthase in Lycopersicon  
 RT esculentum and an anthocyanin-deficient mutant.";  
 RL Mol. Gen. Genet. 224:279-288(1990).  
 CC -1- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-  
 CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)  
 CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO  
 CC NARINGENIN.  
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +  
 CC naringenin-chalcone + 3 CO(2).  
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF  
 CC FLAVONOIDS. A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY  
 CC OF WHICH ARE BRIGHTLY COLORED.  
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X55194; CAA38980.1; -  
 DR PIR: S12223; S12223.  
 DR InterPro: IPR001099; Chal\_stil\_synth.  
 DR Pfam: PF00195; Chal\_stil\_synth; 1.  
 DR Pfam: PF02797; Chal\_stil\_synth; 1.  
 DR PROSITE: PS00441; CHALCONE\_SYNTH; 1.  
 KW Flavonoid biosynthesis; Transferase; Acyltransferase;  
 KW Multigene family.  
 FT ACT\_SITE 164 164 BY SIMILARITY.  
 SQ SEQUENCE 389 AA; 42552 MW; 553DC69E5EA96A8B CRC64;

```

Query Match          91.7%; Score 22; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 gtxxps 6
      || ||
DB      23 GTSTPS 28

RESULT 7
CHS2 LYCES          STANDARD;          PRT;          389 AA.
AC      P23419;
DT      01-NOV-1991 (Rel. 20, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).
DS      CHS2
OS      Lycopersicon esculentum (Tomato).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC      Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
NCBI_TaxID=4081;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Cotyledon, Hypocotyl, and Leaf;
RX      MEDLINE=91117196; PubMed=1980524;
RA      O'Neill S.D., Tong Y., Spoerlein B., Forkmann G., Yoder J.I.;
RT      "Molecular genetic analysis of chalcone synthase in Lycopersicon
RT      esculentum and an anthocyanin-deficient mutant.";
RL      Mol. Gen. Genet. 224:279-288(1990).
CC      -!- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4',2',4',6'-
CC      TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC      WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC      NARINGENIN.
CC      -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC      naringenin-chalcone + 3 CO(2)
CC      -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC      FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC      OF WHICH ARE BRIGHTLY COLORED.
CC      -!- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X55195; CAA38981.1; -
CC      PIR; S12224; S12224.
CC      InterPro; IPR001099; Chal_stil_synt.
CC      Pfam; PF00195; Chal_stil_synt_1
CC      Pfam; PF02797; Chal_stil_syntC; 1.
CC      PROSITE; PS00441; CHALCONE_SYNTH; 1.
CC      Flavonoid biosynthesis; Transferrase; Acyltransferase;
CC      Multigene family.
CC      ACT_SITE 164 164          BY SIMILARITY.
CC      SEQUENCE 389 AA; 42730 MW; F92F46BB3B5FC32F CRC64;

Query Match          91.7%; Score 22; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 gtxxps 6
      || ||
DB      23 GTATPS 28

RESULT 8

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-CV. RED PONTIAC;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Joung Y.-H., Joung H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (In) Plant Gene Register PGR96-027.
CC -1- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U47739; AAB67734.1; -.
DR Mendel; 10624; SOLTU;1193;1.
DR InterPro; IPR001099; Chal_stil_synt.
DR Pfam; PF00195; Chal_stil_synt; 1.
DR Pfam; PF02797; Chal_stil_synt; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
FT ACT_SITE 164
FT SITE 164 BY SIMILARITY
SQ SEQUENCE 389 AA; 42562 MW; B181D9C6B9170A34 CRC64;

Query Match 91.7%; Score 22; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 23 GTSTPS 28

RESULT 10
CHSB_SOLTU STANDARD; PRT; 389 AA.
AC Q43163;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chalcone synthase 1B (EC 2.3.1.74) (Naringenin-chalcone synthase 1B).
GN CHS1B.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. RED PONTIAC;
RA Jeon J.-H., Kim H.-S., Choi K.-H., Joung Y.-H., Joung H., Byun S.-M.;
RT "Characterization of two members of the chalcone synthase gene family
RT from Solanum tuberosum L.";
RL (In) Plant Gene Register PGR96-027.
CC -1- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).

```

```

CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U47740; AAB67735.1; -.
DR Mendel; 10625; SOLTU;1193;2.
DR InterPro; IPR001099; Chal_stil_synt.
DR Pfam; PF00195; Chal_stil_synt; 1.
DR Pfam; PF02797; Chal_stil_synt; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase;
KW Multigene family.
FT ACT_SITE 164
FT SITE 164 BY SIMILARITY
SQ SEQUENCE 389 AA; 42548 MW; E7334A1A34D8D1CC CRC64;

Query Match 91.7%; Score 22; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 23 GTSTPS 28

RESULT 11
CHSJ_PETHY STANDARD; PRT; 389 AA.
AC P22928;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chalcone synthase J (EC 2.3.1.74) (Naringenin-chalcone synthase J).
GN CHSJ.
OS Petunia hybrida (Petunia).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.
OX NCBI_TaxID=4102;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. VIOLET 30; TISSUE=Leaf;
RX MEDLINE=90034197; PubMed=2806915;
RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;
RT "Cloning and molecular characterization of the chalcone synthase
RT multigene family of Petunia hybrida.";
RL Gene 81:245-257(1989).
RN [2]
RP SEQUENCE OF 71-389 FROM N.A.
RC STRAIN-CV. WHITE 137; TISSUE=Anther;
RA van Tunen A.J.;
RL Submitted (MAR-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14597; CAA32737.1; -.
DR EMBL; X14599; CAA32739.1; -.
DR PIR; JS0309; SYPJGJ.
DR InterPro; IPR001099; Chal_stil_synt.
DR Pfam; PF00195; Chal_stil_synt; 1.
DR Pfam; PF02797; Chal_stil_syntC; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
DR Flavonoid biosynthesis; Transferase; Acyltransferase.
KW Multigene family.
FT ACT_SITE 164 BY SIMILARITY.
FT CONFLICT 75 E -> V (IN REF. 2).
SQ SEQUENCE 389 AA; 42558 MW; F2B3CDD82E6FDE7D CRC64;

Query Match 91.7%; Score 22; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 23 GTATPS 28

RESULT 12
CHSY_CATRO STANDARD; PRT; 389 AA.
AC Q92RS4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chalcone synthase (EC 2.3.1.74) (Naringenin-chalcone synthase).
GN CHS.
OS Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Gentianales; Apocynaceae; Rauvolfioideae;
OC Vinceae; Catharanthus.
OX NCBI_TaxID=4058;
RN [1]
RP SEQUENCE FROM N.A.
RA Kaltenbach M., Schroeder G., Schmelzer E., Lutz V., Schroeder J.;
RT "Flavonoid hydroxylase from Catharanthus roseus: cDNA, heterologous
RT expression, enzyme properties, and cell-type specific expression in
RT plants.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -!- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ131813; CAA10511.1; -.
DR Mendel; 36879; Catro; I193; 36879.
```

```
DR InterPro; IPR001099; Chal_stil_synt.
DR Pfam; PF00195; Chal_stil_synt; 1.
DR Pfam; PF02797; Chal_stil_syntC; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase.
FT ACT_SITE 164 BY SIMILARITY.
SQ SEQUENCE 389 AA; 42492 MW; B350742DB4A19186 CRC64;

Query Match 91.7%; Score 22; DB 1; Length 389;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 23 GTSTPS 28

RESULT 13
CHSY_PHAVU STANDARD; PRT; 389 AA.
AC P49440;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chalcone synthase 17 (EC 2.3.1.74) (Naringenin-chalcone synthase 17).
GN CHS17.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CANADIAN WONDER;
RX MEDLINE=88142539; PubMed=3481420;
RA Ryder T.B., Hedrick S.A., Bell J.N., Liang X., Clouse S.D.,
RA Lamb C.J.;
RT "Organization and differential activation of a gene family encoding
RT the plant defense enzyme chalcone synthase in Phaseolus vulgaris.";
RL Mol. Gen. Genet. 210:219-233(1987).
CC -!- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -!- INDUCTION: BY IRRADIATION, ELICITOR, INFECTION OR WOUNDING.
CC -!- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC -----
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X06411; CAA29700.1; -.
DR InterPro; IPR001099; Chal_stil_synt.
DR Pfam; PF00195; Chal_stil_synt; 1.
DR Pfam; PF02797; Chal_stil_syntC; 1.
DR PROSITE; PS00441; CHALCONE_SYNTH; 1.
KW Flavonoid biosynthesis; Transferase; Acyltransferase.
KW Multigene family.
FT ACT_SITE 164 BY SIMILARITY.
SQ SEQUENCE 389 AA; 42654 MW; 25F13C3C4D654F47 CRC64;

Query Match 91.7%; Score 22; DB 1; Length 389;
```

```

Best Local Similarity 66.7%; Pred. No. 92;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 23 GTATPS 28

RESULT 14
Y462_TREPA
ID Y462_TREPA STANDARD; PRT; 392 AA.
AC 083475; 083476;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein TP0462/TP0463 precursor.
GN TP0462/TP0463
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN SEQUENCE FROM N.A.
RP STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Ariach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. REF.1 SEQUENCE DIFFERS
CC FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 279.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001223; AAC65455.1; ALT_FRAME.
DR EMBL; AE001223; AAC65456.1; ALT_FRAME.
DR TIGR; TP0462; -.
DR TIGR; TP0463; -.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 392 HYPOTHETICAL LIPOPROTEIN TP0462/TP0463.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN 148 163 GLY/SER-RICH.
SQ SEQUENCE 392 AA; 40261 MW; F2689260172607AE CRC64;

Query Match 91.7%; Score 22; DB 1; Length 392;
Best Local Similarity 66.7%; Pred. No. 93;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 116 GTTTPS 121

RESULT 15
CHSY_BETVE
ID Y462_TREPA STANDARD; PRT; 392 AA.
AC 083475; 083476;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein TP0462/TP0463 precursor.
GN TP0462/TP0463
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN SEQUENCE FROM N.A.
RP STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Ariach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE TP013X FAMILY OF LIPOPROTEINS.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. REF.1 SEQUENCE DIFFERS
CC FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 279.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001223; AAC65455.1; ALT_FRAME.
DR EMBL; AE001223; AAC65456.1; ALT_FRAME.
DR TIGR; TP0462; -.
DR TIGR; TP0463; -.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Hypothetical protein; Lipoprotein; Membrane; Signal;
KW complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 392 HYPOTHETICAL LIPOPROTEIN TP0462/TP0463.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).
FT DOMAIN 148 163 GLY/SER-RICH.
SQ SEQUENCE 392 AA; 40261 MW; F2689260172607AE CRC64;

Query Match 91.7%; Score 22; DB 1; Length 395;
Best Local Similarity 66.7%; Pred. No. 93;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 23 GTATPS 28

RESULT 16
CHS2_DAUCA
ID CHS2_DAUCA STANDARD; PRT; 397 AA.
AC 092540;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chalcone synthase 2 (EC 2.3.1.74) (Naringenin-chalcone synthase 2).

```

```

GN CHS2.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RA Hirner A.A., Seitz H.U.;
RT "Cloning of two different chalcone synthase isoforms from Daucus
RT carota L. and their differential expression in organs from european
RT wild carrot and in UV-A irradiated cell cultures.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -!- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ206780; CAA07245.1; -.
CC InterPro: IPR001099; Chal_stil_synth.
CC Pfam: PF00195; Chal_stil_synth; 1.
CC ProSITE; PS02797; Chal_stil_synthC; 1.
CC PROSITE; PS00441; CHALCONE_SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase.
CC ACT_SITE 168. 168 BY SIMILARITY.
CC SEQUENCE 397 AA; 43559 MW; 9386F44B9132EDEE CRC64;
CC
CC Query Match 91.7%; Score 22; DB 1; Length 397;
CC Best Local Similarity 66.7%; Pred. No. 94;
CC Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 1 gtxxps 6
CC DB 27 GTATPS 32
CC
CC RESULT 17
CC CHSY_CALCH STANDARD; PRT; 398 AA.
CC AC P48385;
CC DT 01-FEB-1996 (Rel. 33; Created)
CC DT 15-DEC-1998 (Rel. 37; Last sequence update)
CC DT 15-JUL-1999 (Rel. 38; Last annotation update)
CC DE Chalcone synthase (EC 2.3.1.74) (Naringenin-chalcone synthase).
CC CHS.
CC OS Callistephus chinensis (China aster).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
CC Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae;
CC Astereae; Callistephus.
CC OX NCBI_TaxID=13379;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-L 01; TISSUE=Petal;
CC RA Henkel J., Wassenecker M., Sommer H., Forkmann G.;
CC Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

```

```

CC -!- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -!- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z67988; CAA91930.1; -.
CC InterPro: IPR001099; Chal_stil_synth.
CC Pfam: PF00195; Chal_stil_synth; 1.
CC ProSITE; PS02797; Chal_stil_synthC; 1.
CC PROSITE; PS00441; CHALCONE_SYNTH; 1.
CC Flavonoid biosynthesis; Transferase; Acyltransferase.
CC ACT_SITE 167 167 BY SIMILARITY.
CC SEQUENCE 398 AA; 43541 MW; 8B7D4E11B4FAFFC3 CRC64;
CC
CC Query Match 91.7%; Score 22; DB 1; Length 398;
CC Best Local Similarity 66.7%; Pred. No. 94;
CC Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 1 gtxxps 6
CC DB 26 GTATPS 31
CC
CC RESULT 18
CC CHSY_PETCR STANDARD; PRT; 398 AA.
CC AC P16107;
CC DT 01-APR-1990 (Rel. 14; Created)
CC DT 01-APR-1990 (Rel. 14; Last sequence update)
CC DT 15-JUL-1999 (Rel. 38; Last annotation update)
CC DE Chalcone synthase (EC 2.3.1.74) (Naringenin-chalcone synthase).
CC CHS.
CC OS Petroselinum crispum (Parsley) (Petroselinum hortense).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
CC Asteridae; euasterids II; Apiales; Apiaceae; Petroselinum.
CC OX NCBI_TaxID=4043;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RA Reimold U., Kroeger M., Kreuzaler F., Hahlbrock K.;
CC "Coding and 3' non-coding nucleotide sequence of chalcone synthase
CC mRNA and assignment of amino acid sequence of the enzyme.";
CC EMBO J. 2:1801-1805(1983).
CC -!- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-
CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)
CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO
CC NARINGENIN.
CC -!- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +
CC naringenin-chalcone + 3 CO(2).
CC -!- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF
CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY
CC OF WHICH ARE BRIGHTLY COLORED.
CC -!- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; V01538; CAA24779.1; -  
 DR PIR; S42523; S42523.  
 DR InterPro; IPR001099; Chal\_stil\_synth.  
 DR Pfam; PF00195; Chal\_stil\_synth; 1.  
 DR Pfam; PF02797; Chal\_stil\_synth; 1.  
 DR PROSITE; PS00441; CHALCONE\_SYNTH; 1.  
 KW Flavonoid biosynthesis; Transferase; Acyltransferase.  
 FT ACT\_SITE 169 169 BY SIMILARITY.  
 SQ SEQUENCE 398 AA; 43735 MW; E8B19149AD3DABIE CRC64;

Query Match 91.7%; Score 22; DB 1; Length 398;  
 Best Local Similarity 66.7%; Pred. No. 94;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 gtxxps 6  
 || ||  
 Db 28 GTATPS 33

RESULT 19  
 CHSD\_PETHY STANDARD; PRT; 419 AA.  
 AC P22925;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Chalcone synthase D (EC 2.3.1.74) (Naringenin-chalcone synthase D).  
 GN CHSD.  
 OS Petunia hybrida (Petunia).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; euasterids I; Solanales; Solanaceae; Petunia.  
 OX NCBI\_TaxID=4102;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, VIOLET 30; TISSUE=Leaf;  
 RX MEDLINE=90034197; PubMed=2806915;  
 RA Koes R.E., Spelt C.E., van den Elzen P.J.M., Mol J.N.M.;  
 RT "Cloning and molecular characterization of the chalcone synthase  
 RT multigene family of *Petunia hybrida*.";  
 RL Gene 81:245-257(1989).  
 CC -1- FUNCTION: THE PRIMARY PRODUCT OF THIS ENZYME IS 4,2',4',6'-  
 CC TETRAHYDROXYCHALCONE (ALSO TERMED NARINGENIN-CHALCONE OR CHALCONE)  
 CC WHICH CAN UNDER SPECIFIC CONDITIONS SPONTANEOUSLY ISOMERIZE INTO  
 CC NARINGENIN.  
 CC -1- CATALYTIC ACTIVITY: 3 malonyl-CoA + 4-coumaroyl-CoA = 4 CoA +  
 CC naringenin-chalcone + 3 CO(2).  
 CC -1- PATHWAY: PART OF THE BIOSYNTHETIC PATHWAY FOR ALL CLASSES OF  
 CC FLAVONOIDS, A LARGE CLASS OF SECONDARY PLANT METABOLITES, MANY  
 CC OF WHICH ARE BRIGHTLY COLORED.  
 CC -1- SIMILARITY: BELONGS TO THE CHALCONE/STILBENE SYNTHASES FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X14593; CAA32733.1; -  
 DR PIR; JS0312; SYPJCD.  
 DR InterPro; IPR001099; Chal\_stil\_synth.  
 DR Pfam; PF00195; Chal\_stil\_synth; 1.  
 DR Pfam; PF02797; Chal\_stil\_synth; 1.  
 DR PROSITE; PS00441; CHALCONE\_SYNTH; 1.  
 KW Flavonoid biosynthesis; Transferase; Acyltransferase;  
 KW Multigene family.

FT ACT\_SITE 164 164 BY SIMILARITY.  
 SQ SEQUENCE 419 AA; 45979 MW; 55827338DC812ACD CRC64;

Query Match 91.7%; Score 22; DB 1; Length 419;  
 Best Local Similarity 66.7%; Pred. No. 99;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 gtxxps 6  
 || ||  
 Db 23 GTATPS 28

RESULT 20  
 UGTB\_CAEEL STANDARD; PRT; 435 AA.  
 ID UGTB\_CAEEL  
 AC Q22180;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative UDP-glucuronosyltransferase UGT11 (EC 2.4.1.17) (UDPGT).  
 GN UGT11 OR T04H1.7;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 CC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Harris B.;  
 RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: UDP-glucuronate + acceptor = UDP + acceptor  
 CC beta-D-glucuronoside.  
 CC -1- SIMILARITY: BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; Z78200; CAB01584.1; -  
 DR WormPep; T04H1.7; CE13179.  
 DR InterPro; IPR002213; UDPGT.  
 DR Pfam; PF00201; UDPGT; 1.  
 DR PROSITE; PS00375; UDPGT; 1.  
 KW Hypothetical protein; Transferase; Glycosyltransferase;  
 KW Multigene family.  
 SQ SEQUENCE 435 AA; 50020 MW; AF04A0DE7208CBF3 CRC64;

Query Match 91.7%; Score 22; DB 1; Length 435;  
 Best Local Similarity 66.7%; Pred. No. 1e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 || ||  
 Db 300 GTTTPS 305

RESULT 21  
 DNAK\_PAVLU STANDARD; PRT; 629 AA.  
 ID DNAK\_PAVLU  
 AC P30722;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Chaperone protein dnaK (Heat shock protein 70) (Heat shock 70 kDa  
 DE protein) (HSP70).  
 GN DNAK OR HSP70.  
 OS Pavlova lutheri (Monochrysis lutheri).

OG Chloroplast.  
OC Eukaryota; Haptophyceae; Pavlova; Pavlova.  
OX NCBI\_TaxID=2832;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92163012; PubMed=1536924;  
RA Scaramuzzi C.D., Stokes H.W., Hiller R.G.;  
RT "Heat shock Hsp70 protein is chloroplast-encoded in the chromophytic  
RT alga Pavlova lutherii.";  
RL Plant Mol. Biol. 18:467-476(1992).  
RN [5]  
CC -!- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X59555; CAA42154.1; -  
DR PIR: S20516; S20516.  
DR HSP: P04475; LDG4.  
DR InterPro: IPR001023; HSP70.  
DR Pfam: PF00012; HSP70; 1.  
DR PRINTS: PR00301; HEATSHOCK70.  
DR PROSITE: PS00297; HSP70.1; 1.  
DR PROSITE: PS00329; HSP70.2; 1.  
DR PROSITE: PS01036; HSP70.3; 1.  
KW Chaperone; ATP-binding; Heat shock; Chloroplast.  
SQ SEQUENCE 629 AA; 68792 MW; C388D0C369979D66 CRC64;

Query Match 91.7%; Score 22; DB 1; Length 629;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
DB 33 GTTTPS 38

RESULT 22  
SPL\_HUMAN  
ID SPL\_HUMAN STANDARD; PRT; 785 AA.  
AC P08047; Q9NVE7; Q9H3Q5;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transcription factor Spl.  
OS Spl OR TSFPI.  
GN Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE OF 4-785 FROM N.A.  
RC TISSUE=Cervical carcinoma;  
RA Haggart M.H., Ladurner A.G.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-558 FROM N.A.  
RX MEDLINE=20545561; PubMed=10973950;  
RA Takahara T., Kanazu S., Yanagisawa S., Akanuma H.;  
RT "Heterogeneous Spl mRNAs in human HepG2 cells include a product of  
RT homotypic trans-splicing.";  
RL J. Biol. Chem. 275:38067-38072(2000).  
RN [3]  
RP SEQUENCE OF 90-785 FROM N.A.; AND SEQUENCE OF 359-375 AND 670-675.  
RX MEDLINE=8808466; PubMed=3319186;  
RA Kadosawa J.T., Carner K.R., Masiarz F.R., Tjian R.;  
RT "Isolation of cDNA encoding transcription factor Spl and functional

RT analysis of the DNA binding domain.";  
RL Cell 51:1079-1090(1987).  
RN [4]  
RP O-GLYCOSYLATION.  
RX MEDLINE=89003041; PubMed=3139301;  
RA Jackson S.P., Tjian R.;  
RT "O-glycosylation of eukaryotic transcription factors: implications  
RT for mechanisms of transcriptional regulation.";  
RL Cell 55:125-133(1988).  
RN [5]  
RP STRUCTURE BY NMR OF 654-684 AND 684-712.  
RX MEDLINE=97218212; PubMed=9065444;  
RA Narayan V.A., Kriwacki R.W., Caradonna J.P.;  
RT "Structures of zinc finger domains from transcription factor Spl.  
RT Insights into sequence-specific protein-DNA recognition.";  
RL J. Biol. Chem. 272:7801-7809(1997).  
RN [6]  
RP IDENTIFICATION OF SEROTONIN 1A RECEPTOR PROMOTER BINDING SITES.  
RX MEDLINE=96224025; PubMed=8626793;  
RA Parks C.L., Shenk T.;  
RT "The serotonin 1a receptor gene contains a TATA-less promoter that  
RT responds to MAZ and Spl.";  
RL J. Biol. Chem. 271:4417-4430(1996).  
CC -!- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY  
CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL  
CC RECOGNITION SITES. CAN INTERACT WITH G/C-RICH MOTIFS FROM  
CC SEROTONIN RECEPTOR PROMOTER.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- PTM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS.  
CC -!- SIMILARITY: BELONGS TO THE SPL FAMILY OF C2H2-TYPE ZINC-FINGER  
CC PROTEINS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AF252284; AAF67726.1; -  
DR EMBL: AB039286; BAB13476.1; -  
DR EMBL: J03133; AAA61154.1; -  
DR PIR: A29635; A29635.  
DR PDB: 1SP1; 21-APR-97.  
DR PDB: 1SP2; 21-APR-97.  
DR TRANSFAC: T00759; -  
DR GlycoSuiteDB: P08047; -  
DR MIM: 189906; -  
DR InterPro: IPR000822; znf-C2H2.  
DR Pfam: PF00096; zf-C2H2; 3.  
DR SMART: SM00355; znf-C2H2; 3.  
DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 3.  
KW Transcription regulation; Activator; zinc-finger; Metal-binding;  
KW DNA-binding; Nuclear protein; Repeat; Glycoprotein; 3D-structure.  
FT DOMAIN 626 708  
FT ZN\_FING 626 650  
FT ZN\_FING 656 680  
FT ZN\_FING 686 708  
FT ZN\_FING 686 666  
FT CONFLICT 366 366 D -> G (IN REF. 3; AA SEQUENCE).  
FT CONFLICT 670 670 S -> F (IN REF. 3; AA SEQUENCE).  
SQ SEQUENCE 785 AA; 80693 MW; 43893DBF6518B9EA CRC64;

Query Match 91.7%; Score 22; DB 1; Length 785;  
Best Local Similarity 66.7%; Pred. No. 1.8e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
DB 736 GTATPS 741

```
RESULT 23
SPL_RAT
ID SPL_RAT STANDARD; PRT; 788 AA.
AC Q01714;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor Spl.
GN SPL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93010958; PubMed=1356762;
RA Imataka H., Sogawa K., Yasumoto K., Kikuchi Y., Sasano K.,
RA Kobayashi A., Hayami M., Fujii-Kuriyama Y.;
RT "Two regulatory proteins that bind to the basic transcription element
RT (BTE), a GC box sequence in the promoter region of the rat P-4501A1
RT gene.";
RL EMBO J. 11:3663-3671(1992).
CC -!- FUNCTION: BINDS TO GC BOX PROMOTERS ELEMENTS AND SELECTIVELY
CC ACTIVATES MRNA SYNTHESIS FROM GENES THAT CONTAIN FUNCTIONAL
CC RECOGNITION SITES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS
CC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SPL FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; D12768; BAA02235.1; -
DR PIR; J50747; JS0747.
DR HSP; P08047; ISP1.
DR TRANSFAC; T00754; -.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF000096; zf-C2H2; 3.
DR PRINTS; PR00048; ZNCFINGER.
DR SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat; Glycoprotein.
FT DOMAIN 629 711 ZINC FINGERS.
FT ZN_FING 629 653 C2H2-TYPE.
FT ZN_FING 659 683 C2H2-TYPE.
FT ZN_FING 689 711 C2H2-TYPE.
SQ SEQUENCE 788 AA; 81015 MW; AA2B0CAB81AAB80C CRC64;

Query Match 91.7%; Score 22; DB 1; Length 788;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 739 GTATPS 744

RESULT 24
ULK1_HUMAN
ID ULK1_HUMAN STANDARD; PRT; 1050 AA.
AC O75385;

Query Match 91.7%; Score 22; DB 1; Length 1050;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 745 GTSSPS 750

RESULT 25
GNRP_MOUSE
ID GNRP_MOUSE STANDARD; PRT; 1262 AA.
AC P27671;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Guanine nucleotide releasing protein (GNRP) (Ras-specific nucleotide
DE exchange factor CDC25) (CDC25Mm).
GN RASGRF1 OR CDC25 OR GRF1.
```

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C;  
 RX MEDLINE=93010996; PubMed=1396590;  
 RA Cen H., Lowy D.D.;  
 RT "Isolation of multiple mouse cDNAs with coding homology to  
 RT Saccharomyces cerevisiae CDC25: identification of a region related to  
 RT Bcr, Vav, Dbl and CDC24";  
 RL EMBO J. 11:4007-4015(1992).  
 RN [2]  
 RP SEQUENCE OF 791-1262 FROM N.A.  
 RC STRAIN=SWISS; TISSUE=Brain;  
 RX MEDLINE=92289680; PubMed=1376246;  
 RA Martegani E., Vanoni M., Zippel R., Coccetti P., Brambilla R.,  
 RA Ferrarri C., Sturani E.P., Alberghina L.;  
 RT "Cloning by functional complementation of a mouse cDNA encoding a  
 RT homologue of CDC25, a Saccharomyces cerevisiae RAS activator";  
 RL EMBO J. 11:2151-2157(1992).  
 RN [3]  
 RP SEQUENCE OF 1031-1226 FROM N.A.  
 RC MEDLINE=92357779; PubMed=1379731;  
 RX Wei W., Mosteller R.D., Sanyal P., Gonzales E., McKinney D.,  
 RA Dasgupta C., Li P., Liu B.X., Broek D.;  
 RT "Identification of a mammalian gene structurally and functionally  
 RT related to the CDC25 gene of Saccharomyces cerevisiae";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7100-7104(1992).  
 CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
 CC -!- TISSUE SPECIFICITY: BRAIN.  
 CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 1 DEL-HOMOLOGY DOMAIN (DH).  
 CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L20899; AAA02741.1; -;  
 DR EMBL; X59868; CAA42525.1; -;  
 DR PIR; S20730; S20730.  
 DR PIR; S22693; S22693.  
 DR MGI; 99694; Rasgrfl.  
 DR InterPro; IPR001331; GDS\_CDC24.  
 DR InterPro; IPR000048; IQ.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000651; RasGEFN.  
 DR InterPro; IPR001895; RasGEF\_CDC25.  
 DR InterPro; IPR000219; RasGEF.  
 DR Pfam; PF00612; IQ; 1.  
 DR Pfam; PF00169; PH; 2.  
 DR Pfam; PF00617; RasGEF; 1.  
 DR Pfam; PF00618; RasGEFN; 1.  
 DR Pfam; PF00621; RhoGEF; 1.  
 DR SMART; SM00015; IQ; 1.  
 DR SMART; SM00233; PH; 2.  
 DR SMART; SM00147; RasGEF; 1.  
 DR SMART; SM00229; RasGEFN; 2.  
 DR SMART; SM00325; RhoGEF; 1.  
 DR PROSITE; PS50010; DH\_1; 1.  
 DR PROSITE; PS00741; DH\_1; 1.  
 DR PROSITE; PS00720; GDS\_CDC25; 1.  
 DR PROSITE; PS50096; IQ; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; 2.  
 KW Guanine-nucleotide releasing factor; Repeat.  
 FT DOMAIN 22 130

FT DOMAIN 208 233 IQ.  
 FT DOMAIN 244 430 DH.  
 FT DOMAIN 460 588 PH 2.  
 FT DOMAIN 1025 1259 RASGEF.  
 FT CONFLICT 1033 1033 E -> D. (IN REF. 3).  
 SQ SEQUENCE 1262 AA; 144101 MW; 38BFE68F7C228DC8 CRC64;  
 Query Match 91.7%; Score 22; DB 1; Length 1262;  
 Best Local Similarity 66.7%; Pred. No. 2.8e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 gtxxps 6  
 DB 734 GTSSPS 739  
 RESULT 26  
 CID\_DROME  
 ID CID\_DROME STANDARD; PRT; 1377 AA.  
 AC P19538;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Cubitus interruptus dominant protein.  
 GI D.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RX MEDLINE=90346286; PubMed=2166702;  
 RA Orenic T.V., Slusarski D.C., Kroll K.L., Holmgren R.A.;  
 RT "Cloning and characterization of the segment polarity gene cubitus  
 RT interruptus Dominant of Drosophila";  
 RL Genes Dev. 4:1053-1067(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OREGON-R;  
 RX MEDLINE=92146935; PubMed=1686006;  
 RA Berry A.J., Ajioke J.W., Kreitman M.;  
 RT "Lack of polymorphism on the Drosophila fourth chromosome resulting  
 RT from selection";  
 RL Genetics 129:1111-1117(1991).  
 CC -!- FUNCTION: INVOLVED IN SEGMENT POLARITY. IS REQUIRED FOR THE NORMAL  
 CC DEVELOPMENT OF THE POSTERIOR HALF OF EACH EMBRYONIC SEGMENT.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED UNIFORMLY THROUGHOUT THE BLASTODERM  
 CC STAGE AND GASTRULATION AND DOES NOT RESOLVE INTO SEGMENTALLY  
 CC REPEATING STRIPES UNTIL THE END OF THE SHORT PHASE OF GERM-BAND  
 CC EXTENSION.  
 CC -!- SIMILARITY: TO THE GLI-RELATED GROUP OF C2H2-TYPE ZINC-FINGERS  
 CC PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X54360; CAA38244.1; -;  
 DR PIR; A35817; A35817.  
 DR PIR; S12769; S12769.  
 DR HSSP; P08151; 2GLI.  
 DR FlyBase; FBgn0004859; ci.  
 DR InterPro; IPR000822; Znf-C2H2.  
 DR Pfam; PF00096; zf-C2H2; 5.  
 DR PRINTS; PR000048; ZINCINGER.

```
DR SMART; SM00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
KW Developmental protein; Segmentation polarity protein; Zinc-finger;
KW Metal-binding; DNA-binding; Repeat; Nuclear protein.
FT DOMAIN 451 603 ZINC_FINGERS.
FT ZN_FING 451 476 C2H2-TYPE.
FT ZN_FING 484 511 C2H2-TYPE.
FT ZN_FING 517 541 C2H2-TYPE.
FT ZN_FING 547 572 C2H2-TYPE.
FT ZN_FING 578 603 C2H2-TYPE.
SQ SEQUENCE 1377 AA; 150881 MW; 85189314D5E7B3B7 CRC64;

Query Match 91.7%; Score 22; DB 1; Length 1377;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 655 GTSPSPS 660

RESULT 27
ZEPL_HUMAN
ID ZEPL_HUMAN STANDARD; PRT; 2717 AA.
AC P15822;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 40 (Human immunodeficiency virus type I enhancer-
DE binding protein 1) (HIV-1) (Major histocompatibility complex binding
DE protein 1) (MBP-1) (Positive regulatory domain II binding factor 1)
DE (PROII-BF1)
DE
GN HIVP1 OR ZNF40.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90169514; PubMed=2106471;
RA Fan C.M., Maniatis T.;
RT "A DNA-binding protein containing two widely separated zinc finger
RT motifs that recognize the same DNA sequence.";
RL Genes Dev. 4:29-42(1990).
RN [2]
RP STRUCTURE BY NMR OF 2113-2142.
RX MEDLINE=91064333; PubMed=2248949;
RA Omichinski J.G., Clore G.M., Appella E., Sakaguchi K.,
RA Gronenborn A.M.;
RT "High-resolution three-dimensional structure of a single zinc finger
RT from a human enhancer binding protein in solution.";
RL Biochemistry 29:9324-9334(1990).
RN [3]
RP STRUCTURE BY NMR OF 2087-2142.
RX MEDLINE=92232684; PubMed=1567844;
RA Omichinski J.G., Clore G.M., Robien M., Sakaguchi K., Appella E.,
RA Gronenborn A.M.;
RT "High-resolution solution structure of the double Cys2His2 zinc
RT finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917(1992).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGACTTTC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
```

```
CC SMART; SM00355; Znf_C2H2; 5.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51435; CAA35798.1; -
CC PIR; A34203; A34203.
CC PDB; 3ZNF; 15-JAN-92.
CC PDB; 4ZNF; 15-JAN-92.
CC PDB; 1BBO; 31-OCT-93.
CC TRANSFAC; T00497; -
CC MIN; 194540; -
CC InterPro; IPR000822; Znf-C2H2.
CC PRINTS; PF00096; zf-C2H2; 5.
CC PRINTS; PR00048; ZINCFINGER.
CC SMART; SM00355; Znf_C2H2; 4.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
CC Nuclear protein; Repeat; 3D-structure.
CC DOMAIN 406 456 ZINC_FINGERS.
CC ZN_FING 406 428 C2H2-TYPE.
CC ZN_FING 434 456 C2H2-TYPE.
CC DOMAIN 803 806 POLY-SER.
CC ZN_FING 958 981 C2HC-TYPE (POTENTIAL).
CC DOMAIN 2087 2139 ZINC_FINGERS.
CC ZN_FING 2087 2109 C2H2-TYPE.
CC ZN_FING 2115 2139 C2H2-TYPE.
CC STRAND 2088 2088
CC TURN 2090 2092
CC STRAND 2095 2095
CC HELIX 2099 2108
CC TURN 2109 2109
CC STRAND 2115 2116
CC STRAND 2123 2124
CC HELIX 2127 2135
SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 91.7%; Score 22; DB 1; Length 2717;
Best Local Similarity 66.7%; Pred. No. 5.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 185 GTSPSPS 190

RESULT 28
THYG_BOVIN
ID THYG_BOVIN STANDARD; PRT; 2769 AA.
AC P01267; Q18976; Q95478; Q28196;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tyroglobulin precursor.
GN TG.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85296288; PubMed=3855243;
RA Mercken L., Simons M.-J., Swillens S., Massaer M., Vassart G.;
RT "Primary structure of bovine thyroglobulin deduced from the sequence
```

of its 8,431-base complementary DNA.";

[2]

Nature 316:647-651(1985).

RP MEDLINE-85127025; PubMed-3855750;

RA Mercken L., Simons M.-J., de Martynhoff G., Swillens S., Vassart G.;

RT "Presence of homonogenic and repetitive domains in the first 930

RT amino acids of bovine thyroglobulin as deduced from the cDNA

RT sequence.";

RT Eur. J. Biochem. 147:59-64(1985).

[3]

RP MEDLINE-87190432; PubMed-3032624;

RA de Martynhoff G., Pohl V., Mercken L., van Ommen G.-J., Vassart G.;

RT "Structural organization of the bovine thyroglobulin gene and of its

RT 5'-flanking region.";

RT Eur. J. Biochem. 164:591-599(1987).

[4]

RP MEDLINE-1002-1209 FROM N.A.

RA Parma J., Christophe D., Pohl V., Vassart G.;

RT "Structural organization of the 5' region of the thyroglobulin gene.

RT Evidence for intron loss and 'exonization' during evolution.";

RT J. Mol. Biol. 196:769-779(1987).

CC -!- FUNCTION: PRECURSOR OF THE IODINATED THYROID HORMONES THYROXINE

CC (T4) AND TRIIODOTHYRONINE (T3).

CC -!- SUBUNIT: HOMODIMER.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- TISSUE SPECIFICITY: THYROID GLAND SPECIFIC.

CC -!- PTM: SULFATED (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.

CC -!- SIMILARITY: CONTAINS 11 THYROGLOBULIN TYPE-I DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X02815; CAA26584.1; -

DR EMBL; X02155; CAA26090.1; -

DR EMBL; X05380; CAA28971.1; ALT\_SEQ.

DR EMBL; X06071; CAA29457.1; -

DR EMBL; X06072; CAA29457.1; JOINED.

DR EMBL; X06073; CAA29457.1; JOINED.

DR EMBL; X06074; CAA29457.1; JOINED.

DR EMBL; X06075; CAA29457.1; JOINED.

DR PIR; A01533; U1BO.

DR HSSP; P21836; IMAA.

DR InterPro; IPR002018; Carboxylesterase\_B.

DR InterPro; IPR000716; Thyroglobulin\_1.

DR Pfam; PF00135; Coesterase; 2.

DR Pfam; PF00086; thyroglobulin\_1; 8.

DR SMART; SM00211; TV; 10.

DR PROSITE; PS00484; THYROGLOBULIN\_1; 9.

DR PROSITE; PS00941; CARBOXYLESTERASE\_B; 2; 1.

KW Glycoprotein; Repeat; Thyroid hormone; Iodination; Sulfation; Signal.

FT SIGNAL 1 19

FT CHAIN 20 2769

FT DOMAIN 31 92

FT DOMAIN 93 160

FT DOMAIN 161 297

FT DOMAIN 298 358

FT DOMAIN 604 657

FT DOMAIN 661 725

FT DOMAIN 726 921

FT DOMAIN 922 1073

FT DOMAIN 1074 1145

FT DOMAIN 1146 1210

FT REPEAT 1458 1471

FT REPEAT 1472 1488

FT TYPE II.

FT REPEAT 1489 1505

FT DOMAIN 1513 1567

FT REPEAT 1605 1725

FT REPEAT 1726 1893

FT REPEAT 1894 1996

FT REPEAT 1997 2130

FT REPEAT 2131 2188

FT MOD\_RES 24 24

FT MOD\_RES 24 24

FT MOD\_RES 2574 2574

FT MOD\_RES 2588 2588

FT MOD\_RES 2767 2767

FT CARBOHYD 110 110

FT CARBOHYD 198 198

FT CARBOHYD 483 483

FT CARBOHYD 495 495

FT CARBOHYD 747 747

FT CARBOHYD 853 853

FT CARBOHYD 947 947

FT CARBOHYD 1140 1140

FT CARBOHYD 1365 1365

FT CARBOHYD 1776 1776

FT CARBOHYD 1870 1870

FT CARBOHYD 2014 2014

FT CARBOHYD 2123 2123

FT CARBOHYD 2251 2251

FT CARBOHYD 2296 2296

FT CONFLICT 1206 1206

SQ SEQUENCE 2769 AA; 303218 MW; 1C7F227E9101DEZA CRC64;

Query Match 91.7%; Score 22; DB 1; Length 2769;

Best Local Similarity 66.7%; Pred. No. 5.9e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6

DB 2196 GrSSPS 2201

RESULT 29

LYS3\_ECOLI STANDARD; PRT; 45 AA.

ID LYS3\_ECOLI

AC P05821;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Lys3 protein for colicin E1 precursor.

GN Lys.

OS Escherichia coli.

OG Plasmid ColE1.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxId=562;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-86094231; PubMed-3936034;

RA Waleh N.S.; Johnson P.H.;

RT "Structural and functional organization of the colicin E1 operon.";

RL Proc. Natl. Acad. Sci. U.S.A. 82:8389-8393(1985).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-EC31, AND EC71;

RX MEDLINE-95062249; PubMed-7972047;

RA Riley M.A., Tan Y., Wang J.;

RT "Nucleotide polymorphism in colicin E1 and Ia plasmids from natural

RT isolates of Escherichia coli.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:11276-11280(1994).

CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE

CC AND PARTIAL CELL LYSIS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M12543; AAA23067.1; -;  
DR EMBL; U15629; AAA59411.1; -;  
DR EMBL; U15633; AAA59419.1; -;  
DR PIR; B24685; ZHECP1.  
DR Pfam; PF02402; Lysis\_col.  
DR Pfam; PF02402; Lysis\_col; 1.  
DR PRINTS; PR01297; LYSISCOLICIN.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Plasmid; Outer membrane; Lipoprotein; Signal.  
FT SIGNAL 1 17  
FT CHAIN 18 45 LYSIS PROTEIN FOR COLICIN E1.  
FT LIPID 18 18 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 45 AA; 4829 MW; A34D3B4FBC12A13E CRC64;

Query Match 87.5%; Score 21; DB 1; Length 45;  
Best Local Similarity 66.7%; Pred. No. 23;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 29 GTIAPS 34

RESULT 30  
LYS3\_SHISO STANDARD; PRT; 45 AA.  
AC P21185;  
DT 01-MAY-1991 (Rel. 18, Created)  
DT 01-MAY-1991 (Rel. 18, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lysis protein for colicin E1\* precursor.  
GN KIL.  
OS Shigella sonnei.  
OG Plasmid pKY-1.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=624;  
[1]  
RN SEQUENCE FROM N.A.  
RP Higashi M., Hata M., Hase T., Yamaguchi K., Masamune Y.;  
RA "The nucleotide sequence of cea and the region of origin of plasmid  
RT pKY-1.";  
RL J. Gen. Appl. Microbiol. 32:433-442(1986).  
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE  
CC AND PARTIAL CELL LYSIS.  
CC -!- SIMILARITY: 96% IDENTITY TO E. COLI LYSIS PROTEIN PRECURSOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M37218; AAA98158.1; -;  
DR PIR; S10921; S10921.  
DR InterPro; IPR003059; Lysis\_col.  
DR Pfam; PF02402; Lysis\_col; 1.  
DR PRINTS; PR01297; LYSISCOLICIN.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Plasmid; Outer membrane; Lipoprotein; Signal.  
FT SIGNAL 1 17  
FT CHAIN 18 45 LYSIS PROTEIN FOR COLICIN E1\*.  
FT LIPID 18 18 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 45 AA; 4831 MW; A34D232FBC12B99D CRC64;

Query Match 87.5%; Score 21; DB 1; Length 45;  
Best Local Similarity 66.7%; Pred. No. 23;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|| ||  
Db 29 GTVAPS 34

RESULT 31  
LYS2\_ECOLI STANDARD; PRT; 47 AA.  
AC P06963;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lysis protein for colicins E2 and E3 precursor.  
GN HIC OR CELB.  
OS Escherichia coli.  
OG Plasmid Cole2-P9, and Plasmid Cole3-CA38.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
[1]  
RN SEQUENCE FROM N.A.  
RP PLASMID-Cole2-P9;  
RC MEDLINE=86195936; PubMed=3516985;  
RA Toba M., Masaki H., Ohta T.;  
RT "Primary structures of the Cole2-P9 and Cole3-CA38 lysis genes.";  
RL J. Biochem. 99:591-596(1986).  
[2]  
RN SEQUENCE FROM N.A.  
RP PLASMID-Cole2-P9;  
RC MEDLINE=85239907; PubMed=3892228;  
RA Cole S.T., Saint-Joanis B., Pugsley A.P.;  
RT "Molecular characterisation of the colicin E2 operon and  
RT identification of its products.";  
RL Mol. Gen. Genet. 198:465-472(1985).  
[3]  
RN SEQUENCE FROM N.A.  
RP PLASMID-Cole3-CA38;  
RC MEDLINE=85028427; PubMed=6092219;  
RA Watson R.J., Lau P.C.K., Vernet T., Visentin L.P.;  
RT "Characterization and nucleotide sequence of a colicin-release gene  
RT in the hic region of plasmid Cole3-CA38.";  
RL Gene 29:175-184(1984).  
[4]  
RN ERRATUM.  
RP PLASMID-Cole3-CA38;  
RC Watson R.J., Lau P.C.K., Vernet T., Visentin L.P.;  
RA Gene 42:351-353(1986).  
[5]  
RN SEQUENCE OF 1-38 FROM N.A.  
RP PLASMID-Cole3-CA38;  
RC MEDLINE=85210906; PubMed=3889348;  
RA Masaki H., Ohta T.;  
RT "Colicin E3 and its immunity genes.";  
RL J. Mol. Biol. 182:217-227(1985).  
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE  
CC AND PARTIAL CELL LYSIS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X02397; CAA26243.1; -;  
DR EMBL; D00020; BAA00014.1; -;

DR EMBL; D00021; BAA00015.1; -  
DR EMBL; J01574; AAA86419.1; -  
DR EMBL; X03631; CAA27281.1; -  
DR EMBL; X03632; CAA27282.1; -  
DR EMBL; M29885; AAA23070.1; -  
DR PIR; A22383; BVECH8.  
DR PIR; JS0004; JS0004.  
DR PIR; JS0529; JS0529.  
DR InterPro; IPR003059; Lysis\_col.  
DR Pfam; PF02402; Lysis\_col; 1.  
DR PRINTS; PR01297; LYSISCOLICIN.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Plasmid; Outer membrane; Lipoprotein; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 47 LYSIS PROTEIN FOR COLICINS E2 AND E3.  
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 47 AA; 4860 MW; 691E149A8A164A0A CRC64;

Query Match 87.5%; Score 21; DB 1; Length 47;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
DB 31 GTVSPS 36

RESULT 32  
LYS5\_ECOLI  
ID LYS5\_ECOLI STANDARD; PRT; 47 AA.  
AC PL3344;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lysis protein for colicin E5 precursor.  
GN Lys.  
OS Escherichia coli.  
OG Plasmid ColE5-099.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89364708; PubMed=2549375;  
RA Lau P.C.K., Condie J.A.;  
RT "Nucleotide sequences from the colicin E5, E6 and E9 operons:  
RT presence of a degenerate transposon-like structure in the ColE9-J  
RT plasmid.";  
RL Mol. Gen. Genet. 217:269-277(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90218006; PubMed=2561131;  
RA Curtis M.D., James R., Coddington A.;  
RT "An evolutionary relationship between the ColE5-099 and the ColE9-J  
RT plasmids revealed by nucleotide sequencing.";  
RL J. Gen. Microbiol. 135:2783-2788(1989).  
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE  
CC AND PARTIAL CELL LYSIS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X15857; CAA33861.1; -  
DR EMBL; M30445; AAA98053.1; -  
DR PIR; JQ0330; JQ0330.  
DR PIR; C45799; C45799.  
DR InterPro; IPR003059; Lysis\_col.

DR Pfam; PF02402; Lysis\_col; 1.  
DR PRINTS; PR01297; LYSISCOLICIN.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Plasmid; Outer membrane; Lipoprotein; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 47 LYSIS PROTEIN FOR COLICIN E5.  
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 47 AA; 4926 MW; 69133B25CA15A4B8 CRC64;

Query Match 87.5%; Score 21; DB 1; Length 47;  
Best Local Similarity 66.7%; Pred. No. 24;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
DB 31 GTVSPS 36

RESULT 33  
LYS6\_ECOLI  
ID LYS6\_ECOLI STANDARD; PRT; 47 AA.  
AC PL3345;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lysis protein for colicin E6 precursor.  
GN Lys.  
OS Escherichia coli.  
OG Plasmid ColE6-CT14.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89364708; PubMed=2549375;  
RA Lau P.C.K., Condie J.A.;  
RT "Nucleotide sequences from the colicin E5, E6 and E9 operons:  
RT presence of a degenerate transposon-like structure in the ColE9-J  
RT plasmid.";  
RL Mol. Gen. Genet. 217:269-277(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90078082; PubMed=2687234;  
RA Akutsu A., Masaki H., Ohta T.;  
RT "Molecular structure and immunity specificity of colicin E6, an  
RT evolutionary intermediate between E-group colicins and cloacin  
RT DF13.";  
RL J. Bacteriol. 171:6430-6436(1989).  
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE  
CC AND PARTIAL CELL LYSIS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X15856; CAA33858.1; -  
DR EMBL; M31808; AAA23083.1; -  
DR PIR; JQ0328; JQ0328.  
DR PIR; D43716; D43716.  
DR InterPro; IPR003059; Lysis\_col.  
DR Pfam; PF02402; Lysis\_col; 1.  
DR PRINTS; PR01297; LYSISCOLICIN.  
DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
KW Plasmid; Outer membrane; Lipoprotein; Signal.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 47 LYSIS PROTEIN FOR COLICIN E6.  
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).  
SQ SEQUENCE 47 AA; 4902 MW; 691E124F6A057ED8 CRC64;

```
Query Match      87.5%; Score 21; DB 1; Length 47;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 31 GTVSPS 36

RESULT 34
LVS9_ECOLI
ID LVS9_ECOLI STANDARD; PRT; 47 AA.
AC Q03709;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysis protein for colicin E7 precursor.
GN LYS OR CELE7.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91259043; PubMed=2045785;
RA Chak K.F., Kuo W.S., Lu F.M., James R.;
RT "Cloning and characterization of the Cole7 plasmid.";
RL J. Gen. Microbiol. 137:91-100(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-K317;
RA Lau P.C.K., Parsons M.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE
CC AND PARTIAL CELL LYSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57540; AAA23072.1; -.
DR EMBL; X63620; CAA45166.1; -.
DR InterPro; IPR003059; Lysis_col.
DR Pfam; PF02402; Lysis_col; 1.
DR PRINTS; PR01297; LYSISCOLICIN.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Plasmid; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 47 LYSIS PROTEIN FOR COLICIN E7.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 47 AA; 4874 MW; 6900F2A17A057ED8 CRC64;

Query Match      87.5%; Score 21; DB 1; Length 47;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 31 GTVSPS 36

RESULT 35
LVS8_ECOLI
ID LVS8_ECOLI STANDARD; PRT; 47 AA.
AC P10099;

Query Match      87.5%; Score 21; DB 1; Length 47;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 31 GTVSPS 36

RESULT 36
LVS9_ECOLI
ID LVS9_ECOLI STANDARD; PRT; 47 AA.
AC P15176;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysis protein for colicin E9 precursor.
GN LYS.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88121677; PubMed=3323826;
RA Uchimura T., Lau P.C.K.;
RT "Nucleotide sequences from the colicin E8 operon: homology with
RT plasmid Cole2-P9.";
RL Mol. Gen. Genet. 209:489-493(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88257046; PubMed=3290201;
RA Tobu M., Masaki H., Ohta T.;
RT "Colicin E8, a DNase which indicates an evolutionary relationship
RT between colicins E2 and E3.";
RL J. Bacteriol. 170:3237-3242(1988).
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE
CC AND PARTIAL CELL LYSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M21404; AAA23075.1; -.
DR EMBL; X06119; CAA29493.1; -.
DR PIR; C28184; ZHECE8.
DR PIR; S01082; S01082.
DR InterPro; IPR003059; Lysis_col.
DR Pfam; PF02402; Lysis_col; 1.
DR PRINTS; PR01297; LYSISCOLICIN.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Plasmid; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 47 LYSIS PROTEIN FOR COLICIN E8.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 47 AA; 4844 MW; 691E149A8A056B38 CRC64;

Query Match      87.5%; Score 21; DB 1; Length 47;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
Db 31 GTVSPS 36

RESULT 37
LVS9_ECOLI
ID LVS9_ECOLI STANDARD; PRT; 47 AA.
AC P15176;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysis protein for colicin E9 precursor.
GN LYS.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OC NCBI_TaxID=562;
```

```

RN SEQUENCE FROM N.A.
RA Lau P.C.K.;
RL Submitted (DEC-1989) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88034907; PubMed=3312476;
RA James R., Jarvis M., Barker D.E.;
RT "Nucleotide sequence of the immunity and lysis region of the Cole9-J
RT plasmid."
RL J. Gen. Microbiol. 133:1553-1562(1987).
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE
CC AND PARTIAL CELL LYSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X15858; CAA33866.1; -.
CC PIR: M16803; AAA23079.1; -.
CC InterPro: IPR003059; Lysis_col.
CC Pfam: PF02402; Lysis_col; 1.
CC PRINTS: PR01297; LYSISCOLICIN.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Plasmid; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 47 LYSIS PROTEIN FOR COLICIN E9.
FT LIPID 20 20 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 47 AA; 4816 MW; 69062F749A057ED8 CRC64;

Query Match 87.5%; Score 21; DB 1; Length 47;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 31 GTVSPS 36

RESULT 37
LYS0_ECOLI
ID LYS0_ECOLI STANDARD; PRT; 49 AA.
AC P02967;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysis protein precursor (Protein H).
GN H OR CEX.
OS Escherichia coli.
OG Plasmid Clo DF13.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86314306; PubMed=3749334;
RA Nijkamp H.J.J., de Lang R., Stuitje A.R., van den Elsen P.J.M.,
RA Veltkamp E., van Putten A.J.;
RT "The complete nucleotide sequence of the bacteriocinogenic plasmid
RT CloDF13."
RL Plasmid 16:135-160(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=81148852; PubMed=6163089;
RA Stuitje A.R., Speit C.E., Veltkamp E., Nijkamp H.J.J.;
RT "Identification of mutations affecting replication control of plasmid
RT Clo DF13."

```

```

RL Nature 290:264-267(1981).
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE
CC AND PARTIAL CELL LYSIS.
CC -!- MISCELLANEOUS: PLASMID CLO DF13 ORIGINATES FROM ENTEROBACTER
CC CLOACAE BUT IS STABLY MAINTAINED IN AND STUDIED MOSTLY FROM
CC E.COLI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X04466; CAA28145.1; -.
CC PIR: A03515; ZHECP3.
CC InterPro: IPR003059; Lysis_col.
CC Pfam: PF02402; Lysis_col; 1.
CC PRINTS: PR01297; LYSISCOLICIN.
CC PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Plasmid; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 49 LYSIS PROTEIN.
FT LIPID 22 22 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 49 AA; 5157 MW; 1A3DC979EBB0C5DC CRC64;

Query Match 87.5%; Score 21; DB 1; Length 49;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 33 GTVAPS 38

RESULT 38
LYS4_ECOLI
ID LYS4_ECOLI STANDARD; PRT; 52 AA.
AC P09181;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lysis protein for colicin N precursor.
GN CNL.
OS Escherichia coli.
OG Plasmid COLN pCHAP4.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-K12; PubMed=3280946;
RX MEDLINE=88174431; PubMed=3280946;
RA Pugsley A.P.;
RT "The immunity and lysis genes of ColN plasmid pCHAP4."
RL Mol. Gen. Genet. 211:335-341(1988).
CC -!- FUNCTION: LYSIS PROTEINS ARE REQUIRED FOR BOTH COLICIN RELEASE
CC AND PARTIAL CELL LYSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X06933; CAA30019.1; -.
CC PIR: S01760; ZHECNA.
CC InterPro: IPR003059; Lysis_col.
CC Pfam: PF02402; Lysis_col; 1.

```



```
QY 1 gtxxps 6
Db 12 GTETPS 17

RESULT 41
TRH4_ECOLI
ID TRH4_ECOLI STANDARD; PRT; 118 AA.
AC Q00130;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE TRAH protein.
GN TRAH.
OS Escherichia coli.
OG Plasmid Inc-beta RP4.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RC SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.
RP STRAIN=HB101;
RX MEDLINE=92190548; PubMed=1665997;
RA Ziegler G., Pansegrau W., Strack B., Balzer D., Kroeger M.,
RA Kruff V., Lanka E.;
RT "Nucleotide sequence and organization of genes flanking the transfer
RT origin of promiscuous plasmid RP4.";
RL DNA Seq. 1:303-327(1991).
CC -!- FUNCTION: THE INITIATION PROCESS OF TRANSFER DNA SYNTHESIS
CC REQUIRES THE INTERACTION OF AT LEAST THREE PLASMID-SPECIFIC
CC COMPONENTS (TRAH, I, AND J) AT THE TRANSFER ORIGIN RESULTING
CC IN THE ASSEMBLY OF A SPECIALISED NUCLEOPROTEIN COMPLEX - THE
CC RELAXOSOME.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X54459; CAA38335.1; -.
DR PIR; S23000; S23000.
DR Plasmid; Conjugation.
KW INIT.MET 0
FT INIT.MET 0
SQ SEQUENCE 118 AA; 12738 MW; 2FC1445479E6F4B6 CRC64;

Query Match 87.5%; Score 21; DB 1; Length 118;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
Db 37 GTLAPS 42

RESULT 42
RK14_PORPU
ID RK14_PORPU STANDARD; PRT; 122 AA.
AC P51304;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Chloroplast 50S ribosomal protein L14.
GN RPL14.
OS Porphyra purpurea.
OG Chloroplast.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RC SEQUENCE FROM N.A.

Query Match 87.5%; Score 21; DB 1; Length 118;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
Db 37 GTLAPS 42
```

```
RC STRAIN=AVONPORT;
RA Reith M.E., Munholland J.;
RT "Complete nucleotide sequence of the Porphyra purpurea chloroplast
RT genome.";
RL Plant Mol. Biol. Rep. 13:333-335(1995).
CC -!- SIMILARITY: BELONGS TO THE L14P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U38804; AAC08190.1; -.
DR HSSP; P04450; 1WHI.
DR Mendel; 10314; PORPU.rpl14; 1.
DR InterPro; IPR000218; Ribosomal_L14.
DR Pfam; PF00238; Ribosomal_L14; 1.
DR ProDom; PD001093; Ribosomal_L14; 1.
DR PROSITE; PS00049; RIBOSOMAL_L14; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 122 AA; 13411 MW; 49892C2AB0B6BEB2 CRC64;

Query Match 87.5%; Score 21; DB 1; Length 122;
Best Local Similarity 66.7%; Pred. No. 59;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
Db 26 GTSNPS 31

RESULT 43
FMF7_ECOLI
ID FMF7_ECOLI STANDARD; PRT; 170 AA.
AC P25394;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-JUL-1993 (Rel. 26, Last annotation update)
DE F107 fimbrial protein precursor.
GN FEBA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=107/86;
RX MEDLINE=92225617; PubMed=1348723;
RA Imberechts H.A., de Greve H., Schlicker C., Bouchet H., Pohl P.,
RA Charlier G., Vandekerckhove J., van Damme J., van Montagu M.,
RA Lintermans P.;
RT "Characterization of F107 fimbriae of Escherichia coli 107/86, which
RT causes edema disease in pigs, and nucleotide sequence of the F107
RT major fimbrial subunit gene, feba.";
RL Infect. Immun. 60:1963-1971(1992).
CC -!- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -!- SUBCELLULAR LOCATION: Fimbria.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

DR EMBL; M61713; AAA23735.1; -;  
DR EMBL; M61713; AAA23734.1; -;  
DR PIR; A43841;  
DR InterPro: IPR000259; Fimbrial.  
DR Pfam; PF00419; Fimbrial; 1.  
KW Fimbrial; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 170 F107 FIMBRIAL PROTEIN.  
FT DISULFID 37 78 PROBABLE.  
FT SEQUENCE 170 AA; 17329 MW; AF40947CE387692F CRC64;  
SQ  
Query Match 87.5%; Score 21; DB 1; Length 170;  
Best Local Similarity 66.7%; Pred. No. 80;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gtxxps 6  
DB 54 GTVAPS 59  
RESULT 44  
OM24\_ARATH  
ID OM24\_ARATH STANDARD; PRT; 187 AA.  
AC P82805;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mitochondrial import receptor subunit TOM20-4 (Translocase of outer  
DE membrane 20 kDa subunit 4).  
GN TOM20-4 OR AT5G40930 OR MMG1.2.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLUMBIA;  
RX MEDLINE=21016721; PubMed=11130714;  
RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
RA Nakazaki N., Naruo K., Okumura S., Shinpo S., Takeuchi C., Wada T.,  
RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,  
RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,  
RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,  
RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,  
RA Martienssen R., McCombe W.R., Wilson R.K., Murphy G., Bancroft I.,  
RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
RA Entian K.-D., Terryn N., Hartley N., Bent E., Johnson S.,  
RA Langham S.A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,  
RA Ramsperger U., Wedler K., Balke K., Wedler E., Peters S.,  
RA van Staveren M., Dirks W., Mooijman P., Klein Lankhorst R.,  
RA Weitzenger T., Bothe G., Rose M., Hauf J., Bernelsner S., Hempel S.,  
RA Feldpausch M., Lamberth S., Villarriuel R., Gielen J., Ardiles W.,  
RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
RA Schueller C., Zaccaria P., Meves H.-W., Bevan M., Fransch P.F.;  
RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
RT thaliana.";  
RL Nature 408:823-826(2000).  
RN [2]  
RP SEQUENCE OF 102-117.  
RC STRAIN=CV. COLUMBIA;  
RX PubMed=11161051;  
RA Werhahn W., Nlemeyer A., Jaensch L., Kruff V., Schmitz U.K.,  
RA Braun H.-P.;

"Purification and characterization of the preprotein translocase of  
the outer mitochondrial membrane from Arabidopsis thaliana.  
Plant Physiol. 125:943-954(2001).  
-!- FUNCTION: CENTRAL COMPONENT OF THE RECEPTOR COMPLEX RESPONSIBLE  
FOR THE RECOGNITION AND TRANSLLOCATION OF CYTOSOLICALLY SYNTHESIZED  
MITOCHONDRIAL PREPROTEINS. TOGETHER WITH TOM22 FUNCTIONS AS THE  
TRANSIT PEPTIDE RECEPTOR AT THE SURFACE OF THE MITOCHONDRION  
OUTER MEMBRANE AND FACILITATES THE MOVEMENT OF PREPROTEINS INTO  
THE TRANSLLOCATION PORE.  
-!- SUBUNIT: FORMS PART OF MITOCHONDRIAL RECEPTOR COMPLEX.  
-!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
outer membrane.  
-!- PTM: THE N-TERMINUS IS BLOCKED.  
-!- MISCELLANEOUS: THERE ARE FOUR GENES (TOM20-1, TOM20-2, TOM20-3 AND  
TOM20-4) WHICH ENCODE MITOCHONDRIAL IMPORT RECEPTOR SUBUNITS  
TOM20.  
-!- SIMILARITY: BELONGS TO THE TOM20 FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AB023040; BAB10523.1; -;  
KW Transport; Protein transport; Outer membrane; Mitochondrion;  
KW Transmembrane.  
FT DOMAIN 1 160 INTERMEMBRANE (POTENTIAL).  
FT TRANSMEM 161 178 POTENTIAL.  
FT DOMAIN 179 187 CYTOPLASMIC (POTENTIAL).  
FT SEQUENCE 187 AA; 20973 MW; 156DF3D231EA286C CRC64;  
SQ  
Query Match 87.5%; Score 21; DB 1; Length 187;  
Best Local Similarity 66.7%; Pred. No. 88;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 gtxxps 6  
DB 137 GTAGPS 142  
RESULT 45  
Y319\_MYCPN  
ID Y319\_MYCPN STANDARD; PRT; 193 AA.  
AC P75329;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MG319 homolog (H08\_orf193).  
GN MPN454 OR MP387.  
OS Mycoplasma pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OC NCBI\_TaxID=2104;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 29342 / M129;  
RX MEDLINE=97105885; PubMed=8948633;  
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,  
RA Herrmann R.;  
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
RT pneumoniae.";  
RL Nucleic Acids Res. 24:4420-4449(1996).  
RN [2]  
RP This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE000037; AAB96035.1; -  
DR Hypothetical protein; Transmembrane; Complete proteome.  
KW TRANSMEM 10 30 POTENTIAL.  
FT TRANSMEM 136 156 POTENTIAL.  
SQ SEQUENCE 193 AA; 21443 MW; BAE6806C6D80C1D CRC64;  
  
Query Match 87.5%; Score 21; DB 1; Length 193;  
Best Local Similarity 66.7%; Pred. No. 91;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 gtxxps 6  
DB 31 GIVSPS 36  
|| ||  
  
RESULT 46  
COX3\_MYCTU STANDARD; PRT; 203 AA.  
AC Q10385;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable cytochrome c oxidase polypeptide III (EC 1.9.3.1) (Cytochrome  
DE A43 subunit 3).  
GN CTA6 OR RV2193 OR MT2249 OR MTCY190.04.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holtroft S.,  
RA Horsbury T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skellton S., Squares R., Squares R.,  
RA Stulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome  
CC c + 2 H(2)O.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SUBCELLULAR LOCATION: THE CYTOCHROME C OXIDASE SUBUNIT 3 FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z70283; CAA94262.1; -  
DR

DR EMBL; AE007071; AAK46535.1; -  
DR HSP; P00415; LOCC.  
DR TIGR; MT2249; -  
DR Tuberculin; RV2193; -  
DR InterPro; IPR000298; CytC\_oxdse\_III.  
DR Pfam; PF00510; COX3; 1.  
DR PROSITE; PS02553; COX3; 1.  
KW Oxidoreductase; Transmembrane; Complete proteome.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 71 91 POTENTIAL.  
FT TRANSMEM 96 116 POTENTIAL.  
FT TRANSMEM 143 163 POTENTIAL.  
FT TRANSMEM 179 199 POTENTIAL.  
SQ SEQUENCE 203 AA; 22420 MW; 70145380A05BD0C2 CRC64;  
  
Query Match 87.5%; Score 21; DB 1; Length 203;  
Best Local Similarity 66.7%; Pred. No. 95;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 gtxxps 6  
DB 128 GTSIPS 133  
|| ||  
  
RESULT 47  
NODB\_RHISN STANDARD; PRT; 215 AA.  
ID NODB\_RHISN STANDARD; PRT; 215 AA.  
AC P50355;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Chitoooligosaccharide deacetylase (EC 3.5.1.-) (Modulation protein B).  
GN NODB OR Y4HH.  
OS Rhizobium sp. (strain NGR234).  
OG Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97305956; PubMed=9163424;  
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
RA Perret X.;  
RT "Molecular basis of symbiosis between Rhizobium and legumes."  
RL Nature 387:394-401 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95075295; PubMed=7984092;  
RA Relic B., Perret X., Estrada-Garcia M.T., Kocinska J., Golinowski W.,  
RA Krishnan H.B., Pueppke S.G., Broughton W.J.;  
RT "Mod factors of Rhizobium are a key to the legume door."  
RL Mol. Microbiol. 13:171-178 (1994).  
CC -!- FUNCTION: IS INVOLVED IN GENERATING A SMALL HEAT-STABLE COMPOUND  
CC (NOD), AN ACTIVATED OLIGOMER OF N-ACETYLGLUCOSAMINE, THAT  
CC STIMULATES MITOSIS IN VARIOUS PLANT PROTOPLASTS.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: TO OTHER POLYSACCHARIDE DEACETYLASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X73362; CAA51773.1; -  
DR EMBL; AE000076; AAB91696.1; -  
DR InterPro; IPR002509; Polysac\_deacet.  
DR Pfam; PF01522; Polysac\_deacet; 1.  
KW Hydrolase; Nodulation; Plasmid.  
FT CONFLICT 148 149 RP -> S (IN REF. 2).  
DR

SO SEQUENCE 215 AA; 23625 MW; 8A91E2B4E9D6D57F CRC64;

Query Match 87.5%; Score 21; DB 1; Length 215;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 14 GTGAPS 19

RESULT 48

ID XYN2\_ASPNG STANDARD; PRT; 225 AA.  
AC P55330; 012557;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Endo-1,4-beta-xylanase II precursor (EC 3.2.1.8) (Xylanase II)  
DE (1,4-beta-D-xylan xylanohydrolase II).  
GN XYNB.

OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

OX NCBI\_TaxID=5061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IFO 4066;

RA Ito K.;  
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic linkages in xylans.

CC -1- PATHWAY: XLAN DEGRADATION.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY G (FAMILY 11 OF GLYCOSYL HYDROLASES).

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D38071; BAA07265.1; --  
DR HSSP; P09850; 1XNB.

DR InterPro; IPR001137; Glyco\_hydro\_11.

DR Pfam; PF00457; Glyco\_hydro\_11; 1.

DR PRINTS; PR00911; GLHYDLASE11.

DR PROSITE; PS00776; GLYCOSYL\_HYDROL\_F11\_1; 1.

DR PROSITE; PS00777; GLYCOSYL\_HYDROL\_F11\_2; 1.

KW Xylan degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 18 POTENTIAL.

FT PROPEP 19 37

FT CHAIN 38 225 ENDO-1,4-BETA-XYLANASE II.

FT ACT\_SITE 121 121 NUCLEOPHILE (BY SIMILARITY).

FT ACT\_SITE 212 212 PROTON DONOR (BY SIMILARITY).

SQ SEQUENCE 225 AA; 24057 MW; C4B8BB007AB2B8FD CRC64;

Query Match 87.5%; Score 21; DB 1; Length 225;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 99 GTFTPS 104

RESULT 49

PDXJ\_NEIMA STANDARD; PRT; 242 AA.  
ID PDXJ\_NEIMA

AC Q9RQV9;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pyridoxal phosphate biosynthetic protein pdxJ.  
GN PDXJ OR NMA2037.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;

RA Linz B., Schenker M., Achtman M.;

RA "Frequent horizontal genetic exchange between Neisseria meningitidis

RT and commensal neisseriae";

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;

RX MEDLINE=20222556; PubMed=10761919;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,

RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,

RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,

RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

RA Whitehead S., Spratt B.G., Barrell B.G.;

RT "Complete DNA sequence of a serogroup A strain of Neisseria

RL meningitidis Z2491";

RC Nature 404:502-506(2000).

CC -1- FUNCTION: CATALYZES THE CONDENSATION OF 1-DEOXY-D-XYLULOSE-5-

CC PHOSPHATE (DXP) AND 1-AMINO-3-OXO-4-(PHOSPHOHYDROXY)PROPAN-2-ONE

CC TO FORM PYRIDOXINE 5'-PHOSPHATE (PNP) (BY SIMILARITY).

CC -1- PATHWAY: DE NOVO SYNTHESIS OF PYRIDOXINE (VITAMIN B6) AND

CC PYRIDOXAL PHOSPHATE.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- SIMILARITY: BELONGS TO THE PDXJ FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF058689; AAF06688.1; --

DR EMBL; AL162757; CAB85256.1; --

KW Pyridoxine biosynthesis; Complete proteome.

SQ SEQUENCE 242 AA; 26669 MW; 07AC9E95DDA8D090 CRC64;

Query Match 87.5%; Score 21; DB 1; Length 242;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 18 GTTTPS 23

RESULT 50

PDXJ\_NEIMB STANDARD; PRT; 242 AA.

ID PDXJ\_NEIMB

AC Q9K0V9;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Pyridoxal phosphate biosynthetic protein pdxJ.

GN PDXJ OR NMB0448.

OS Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI\_TaxID=491;

RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / SEROGROUP B;  
RX MEDLINE=201755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,  
RA Cottrell M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Masignani V., Pizzi M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain  
MC58.";  
RL Science 287:1809-1815(2000).  
CC -!- FUNCTION: CATALYZES THE CONDENSATION OF 1-DEOXY-D-XYLULOSE-5-  
CC PHOSPHATE (DXP) AND 1-AMINO-3-OXO-4-(PHOSPHOHYDROXY)PROPAN-2-ONE  
CC TO FORM PYRIDOXINE 5'-PHOSPHATE (PNP) (BY SIMILARITY).  
CC -!- PATHWAY: DE NOVO SYNTHESIS OF PYRIDOXINE (VITAMIN B6) AND  
CC PYRIDOXAL PHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
CC -!- SIMILARITY: BELONGS TO THE PDJX FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE002401; AAF40885.1; -.  
DR TIGR; NMB0448; -.  
KW Pyridoxine biosynthesis; Complete proteome.  
SQ SEQUENCE 242 AA; 26565 MW; 5DA0476728AA1485 CRC64;

Query Match 87.5%; Score 21; DB 1; Length 242;  
Best Local Similarity 66.7%; Pred. NO. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 qtxxps 6  
Db 18 GTTYPs 23

Search completed: September 3, 2002, 09:29:46  
Job time: 228 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2002, 09:25:33 ; Search time 25.06 Seconds  
(without alignments)  
41.419 Million cell updates/sec

Title: BASK-853-CLAIM5

Perfect score: 24

Sequence: 1 gtxxps 6

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 50 summaries

Database :

SPTREMBL\_19.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	22	91.7	71	15	Q9IF05 human immun
2	22	91.7	75	10	Q41118 phaseolus v
3	22	91.7	80	15	Q90QH8 human immun
4	22	91.7	89	16	Q9HWZ4 pseudomonas
5	22	91.7	92	15	Q38158 human immun
6	22	91.7	93	15	Q38175 human immun
7	22	91.7	93	15	Q38177 human immun
8	22	91.7	96	15	Q38114 human immun
9	22	91.7	96	15	Q38161 human immun
10	22	91.7	96	15	Q38178 human immun
11	22	91.7	99	15	Q38030 human immun
12	22	91.7	99	15	Q38362 human immun
13	22	91.7	99	15	Q38371 human immun
14	22	91.7	99	15	Q38372 human immun
15	22	91.7	101	15	Q38113 human immun
16	22	91.7	101	15	Q38128 human immun

17	22	91.7	101	15	Q38130	Q38130 human immun
18	22	91.7	101	15	Q38133	Q38133 human immun
19	22	91.7	101	15	Q38138	Q38138 human immun
20	22	91.7	101	15	Q38141	Q38141 human immun
21	22	91.7	105	17	Q9YA39	Q9YA39 aeropyrum p
22	22	91.7	111	15	Q38165	Q38165 human immun
23	22	91.7	111	15	Q38166	Q38166 human immun
24	22	91.7	111	15	Q38169	Q38169 human immun
25	22	91.7	111	15	Q38171	Q38171 human immun
26	22	91.7	112	15	Q75698	Q75698 human immun
27	22	91.7	135	11	Q64053	Q64053 mus sp. orf
28	22	91.7	142	11	Q9D5C6	Q9D5C6 mus musculu
29	22	91.7	156	4	Q9HBN7	Q9HBN7 homo sapien
30	22	91.7	157	4	Q9NWG3	Q9NWG3 homo sapien
31	22	91.7	175	5	Q9VSO7	Q9VSO7 drosophila
32	22	91.7	177	12	Q41995	Q41995 maize rayad
33	22	91.7	178	2	Q9KY82	Q9KY82 streptomyce
34	22	91.7	182	5	Q9VSO6	Q9VSO6 drosophila
35	22	91.7	186	4	Q9BZ94	Q9BZ94 homo sapien
36	22	91.7	196	10	Q82664	Q82664 arabidopsis
37	22	91.7	200	2	Q9F5I3	Q9F5I3 agrobacteri
38	22	91.7	202	16	Q9KOL1	Q9KOL1 vibrio chol
39	22	91.7	218	11	Q9JLH5	Q9JLH5 rattus norv
40	22	91.7	228	2	Q9RDK2	Q9RDK2 streptomyce
41	22	91.7	233	2	Q9KXV0	Q9KXV0 streptomyce
42	22	91.7	237	6	P79289	P79289 sus scrofa
43	22	91.7	247	2	Q9L203	Q9L203 streptomyce
44	22	91.7	253	2	Q9L126	Q9L126 streptomyce
45	22	91.7	253	4	Q43563	Q43563 homo sapien
46	22	91.7	254	16	Q9I170	Q9I170 pseudomonas
47	22	91.7	275	13	Q98TV7	Q98TV7 brachydanio
48	22	91.7	285	10	Q9M0A9	Q9M0A9 arabidopsis
49	22	91.7	288	2	Q9RNC9	Q9RNC9 bartonella
50	22	91.7	288	12	Q9DWD6	Q9DWD6 rat cytomeg

## ALIGNMENTS

```

RESULT 1
ID Q9IF05 PRELIMINARY; PRT; 71 AA.
AC Q9IF05;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1079;
RX MEDLINE=21443958; PubMed=11559796;
RA Peters S., Munoz M., Verly S., Sanchez-Merino V., Lopez-Galindez C.,
RT "Resistance to Nucleoside Analog Reverse Transcriptase Inhibitors
RL Mediated by Human Immunodeficiency Virus Type 1 p6 Protein.";
DR J. Virol. 75:9644-9653(2001).
FT EMBL; AF282969; AAF87830.1; -.
SQ SEQUENCE 71 AA; 7870 MW; 033C600C7F7FBCF3 CRC64;

```

Query Match 91.7%; Score 22; DB 15; Length 71;  
Best Local Similarity 66.7%; Pred. No. le+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
| | |  
Db 39 STATES 44

RESULT 2  
Q41118  
ID Q41118 PRELIMINARY; PRT; 75 AA.  
AC Q41118  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE CHALCONE SYNTHASE MRNA (FRAGMENT).  
OS Phaseolus vulgaris (Kidney bean) (French bean).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.  
OX NCBI\_TaxID=3885;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ryder T.B., Cramer C.L., Bell J.N., Robbins M.P., Dixon R.A.,  
RA Lamb C.J.;  
RT "Elicitor rapidly induces chalcone synthase mRNA in Phaseolus vulgaris  
RT cells at the onset of the phytoalexin defense response.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:5724-5728(1984).  
DR EMBL; K02953; AAA33758.1; -.  
DR InterPro: IPR001099; Chal stil synt.  
DR Pfam: PF00195; Chal stil synt; 1.  
FT NON\_TER 1  
FT NON\_TER 75  
SQ SEQUENCE 75 AA; 8800 MW; 243CBED889A4F207 CRC64;

Query Match 91.7%; Score 22; DB 10; Length 75;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtxxps 6  
|| ||  
Db 18 GTATPS 23

RESULT 3  
Q90QH8  
ID Q90QH8 PRELIMINARY; PRT; 80 AA.  
AC Q90QH8  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE GAG PROTEIN (FRAGMENT).  
GN GAG.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21248068; PubMed=11350662;  
RA Kaufmann G.R., Suzuki K., Cunningham P., Mukaide M., Kondo M.,  
RA Imai M., Zaunders J., Cooper D.A.;  
RT "Impact of HIV Type 1 Protease, Reverse Transcriptase, Cleavage Site,  
RT and p6 Mutations on the Virological Response to Quadruple Therapy with  
RT Saquinavir, Ritonavir, and Two Nucleoside Analogs.";  
RL AIDS Res. Hum. Retroviruses 17:487-497(2001).  
DR EMBL; AF322133; AAK66674.1; -.  
FT NON\_TER 1  
FT NON\_TER 80  
SQ SEQUENCE 80 AA; 8942 MW; 72C5C3148E51E65D CRC64;

Query Match 91.7%; Score 22; DB 15; Length 80;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtxxps 6  
|| ||  
Db 48 GTTTPS 53

RESULT 4  
Q9HWZ4  
ID Q9HWZ4 PRELIMINARY; PRT; 89 AA.  
AC Q9HWZ4  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE HYPOTHETICAL PROTEIN PA4033.  
GN PA4033.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004820; AAG07420.1; -.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 89 AA; 9452 MW; BFD4EBF98897E7E4 CRC64;

Query Match 91.7%; Score 22; DB 16; Length 89;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gtxxps 6  
|| ||  
Db 48 GTASPS 53

RESULT 5  
O38158  
ID O38158 PRELIMINARY; PRT; 92 AA.  
AC O38158  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE GAG PROTEIN (FRAGMENT).  
GN GAG.  
OS Human immunodeficiency virus type 1.  
OC Viruses; Retroviridae; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PATIENT 4;  
RX MEDLINE=97404676; PubMed=9261388;  
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
RA Vasudevachari M.B., Salzman N.P.;  
RT "Drug resistance during indinavir therapy is caused by mutations in  
RT the protease gene and in its Gag substrate cleavage sites.";  
RL J. Virol. 71:6662-6670(1997).  
DR EMBL; AF024054; AAB83833.1; -.  
DR HSSP; P05888; 1AAF.  
DR InterPro: IPR001878; Znf.CCHC.  
DR Pfam: PF00098; Zf-CCHC; 1.  
DR SMART; SM00343; Znf.C2HC; 2.  
KW Zinc-finger.  
FT NON\_TER 1  
FT NON\_TER 92  
SQ SEQUENCE 92 AA; 10403 MW; 27E1160C29F611EA CRC64;

Query Match 91.7%; Score 22; DB 15; Length 92;

```
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
  || ||
Db 72 GTTTPS 77

RESULT 6
O38175 O38175 PRELIMINARY; PRT; 93 AA.
AC O38175;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during Indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024071; AAB83850.1; -.
DR HSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1 1
FT NON_TER 93 93
SQ SEQUENCE 93 AA; 10402 MW; 0E9FF4DC4129C73 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 93;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
  || ||
Db 73 GTATPS 78

RESULT 7
O38177 O38177 PRELIMINARY; PRT; 93 AA.
AC O38177;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during Indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024074; AAB83852.1; -.
DR HSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.

DR Pfam; PF00098; zf-CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1 1
FT NON_TER 93 93
SQ SEQUENCE 93 AA; 10402 MW; 0E9FF4DC4129C73 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 93;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
  || ||
Db 73 GTATPS 78

RESULT 8
O38114 O38114 PRELIMINARY; PRT; 96 AA.
AC O38114;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during Indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024010; AAB83789.1; -.
DR HSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1 1
FT NON_TER 96 96
SQ SEQUENCE 96 AA; 10766 MW; 74A041FFD4FE3935 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 96;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
  || ||
Db 76 GTATPS 81

RESULT 9
O38161 O38161 PRELIMINARY; PRT; 96 AA.
AC O38161;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
```

```
DR Pfam; PF00098; zf-CCHC; 1.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1 1
FT NON_TER 93 93
SQ SEQUENCE 93 AA; 10502 MW; 5615F69550DDA09D CRC64;

Query Match 91.7%; Score 22; DB 15; Length 93;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
  || ||
Db 73 GTATPS 78

RESULT 9
O38161 O38161 PRELIMINARY; PRT; 96 AA.
AC O38161;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
```

```

RX MEDLINE-97404676; PubMed-9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024057; AAB83836.1; -.
DR HSSP; P05888; 1A4F.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 96
FT NON_TER 96
SQ SEQUENCE 96 AA; 10824 MW; 74B74DFE73E93935 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 96;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
|| ||
Db 76 GTTTPS 81

RESULT 10
O38178 PRELIMINARY; PRT; 96 AA.
AC O38178;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE-97404676; PubMed-9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024057; AAB83853.1; -.
DR HSSP; P05888; 1A4F.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 96
FT NON_TER 96
SQ SEQUENCE 96 AA; 10794 MW; 74B756FFC3E93935 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 96;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
|| ||
Db 76 GTTTPS 81

RESULT 11
O38030 PRELIMINARY; PRT; 99 AA.
AC O38030;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

```

```

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 3;
RX MEDLINE-97404676; PubMed-9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF023926; AAB83860.1; -.
DR HSSP; P05888; 1A4F.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 99
FT NON_TER 99
SQ SEQUENCE 99 AA; 11140 MW; 92348443928FDCC0 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 99;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
|| ||
Db 79 GTTTPS 84

RESULT 12
O38362 PRELIMINARY; PRT; 99 AA.
AC O38362;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 7;
RX MEDLINE-97404676; PubMed-9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024261; AAB83135.1; -.
DR HSSP; P05888; 1A4F.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 99
FT NON_TER 99
SQ SEQUENCE 99 AA; 11063 MW; 9818C494BED11470 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 99;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
|| ||

```

```
Db 79 GTATPS 84

RESULT 13
O38371 PRELIMINARY; PRT; 99 AA.
AC O38371;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PATIENT 7;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024270; AAB83144.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11092 MW; 9818C494BD1E1470 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 99;
Best Local Similarity 66.7%; Pred. NO. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
|||
Db 79 GTATPS 84

RESULT 14
O38372 PRELIMINARY; PRT; 99 AA.
AC O38372;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PATIENT 7;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024271; AAB83145.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11092 MW; 9818C494BD1E1470 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 99;
Best Local Similarity 66.7%; Pred. NO. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
|||
Db 79 GTATPS 84

RESULT 15
O38113 PRELIMINARY; PRT; 101 AA.
AC O38113;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024009; AAB83788.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 101
SQ SEQUENCE 101 AA; 11408 MW; 2EF1A3E5CF282854 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 101;
Best Local Similarity 66.7%; Pred. NO. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
|||
Db 81 GTTTPS 86

RESULT 16
O38128 PRELIMINARY; PRT; 101 AA.
AC O38128;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024271; AAB83145.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1
FT NON_TER 101
SQ SEQUENCE 101 AA; 11408 MW; 2EF1A3E5CF282854 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 101;
Best Local Similarity 66.7%; Pred. NO. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6
|||
Db 81 GTTTPS 86
```

RL J. Virol. 71:6662-6670(1997).  
DR EMBL; AF024024; AAB83803.1; -.  
DR HSSP; P05888; IAAF.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF000098; zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
KW Zinc-finger.  
FT NON\_TER 1  
FT NON\_TER 101  
SQ SEQUENCE 101 AA; 11295 MW; EBD5B3278BF1D5FA CRC64;

Query Match 91.7%; Score 22; DB 15; Length 101;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|||  
Db 81 GTTTPS 86

RESULT 17  
O38130 PRELIMINARY; PRT; 101 AA.  
AC O38130;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE GAG PROTEIN (FRAGMENT).  
GN GAG.

OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]

RC STRAIN-PATIENT 4;  
RX MEDLINE=97404676; PubMed=9261388;  
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
RA Vasudevachari M.B., Salzman N.P.;  
RT "Drug resistance during indinavir therapy is caused by mutations in  
the protease gene and in its Gag substrate cleavage sites.";  
RL J. Virol. 71:6662-6670(1997).

DR EMBL; AF024026; AAB83805.1; -.  
DR HSSP; P05888; IAAF.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF000098; zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
KW Zinc-finger.  
FT NON\_TER 1  
FT NON\_TER 101  
SQ SEQUENCE 101 AA; 11408 MW; F7B36D60472BDE0 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 101;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|||  
Db 81 GTTTPS 86

RESULT 18  
O38133 PRELIMINARY; PRT; 101 AA.  
AC O38133;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE GAG PROTEIN (FRAGMENT).  
GN GAG.

OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.

OX NCBI\_TaxID=11676;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 4;  
RX MEDLINE=97404676; PubMed=9261388;  
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
RA Vasudevachari M.B., Salzman N.P.;  
RT "Drug resistance during indinavir therapy is caused by mutations in  
the protease gene and in its Gag substrate cleavage sites.";  
RL J. Virol. 71:6662-6670(1997).  
DR EMBL; AF024029; AAB83808.1; -.  
DR HSSP; P05888; IAAF.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF000098; zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
KW Zinc-finger.  
FT NON\_TER 1  
FT NON\_TER 101  
SQ SEQUENCE 101 AA; 11364 MW; F7B9D83A5781D5FA CRC64;

Query Match 91.7%; Score 22; DB 15; Length 101;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|||  
Db 81 GTATPS 86

RESULT 19  
O38138 PRELIMINARY; PRT; 101 AA.  
AC O38138;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE GAG PROTEIN (FRAGMENT).  
GN GAG.

OS Human immunodeficiency virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN-PATIENT 4;  
RX MEDLINE=97404676; PubMed=9261388;  
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,  
RA Vasudevachari M.B., Salzman N.P.;  
RT "Drug resistance during indinavir therapy is caused by mutations in  
the protease gene and in its Gag substrate cleavage sites.";  
RL J. Virol. 71:6662-6670(1997).

DR EMBL; AF024034; AAB83813.1; -.  
DR HSSP; P05888; IAAF.  
DR InterPro; IPR001878; Znf\_CCHC.  
DR Pfam; PF000098; zf-CCHC; 2.  
DR SMART; SM00343; Znf\_C2HC; 2.  
KW Zinc-finger.  
FT NON\_TER 1  
FT NON\_TER 101  
SQ SEQUENCE 101 AA; 11478 MW; 02BB523A5781CEFB CRC64;

Query Match 91.7%; Score 22; DB 15; Length 101;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
|||  
Db 81 GTATPS 86

RESULT 20  
O38141

```

ID O38141 PRELIMINARY; PRT; 101 AA.
AC O38141;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN HAN.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024037; AAB83816.1; -.
DR HSSP; P05888; IAAF.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1 101
FT NON_TER 101 101
SQ SEQUENCE 101 AA; 11379 MW; F7BB523A5781D5FA CRC64;

Query Match 91.7%; Score 22; DB 15; Length 101;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 81 GTATPS 86

RESULT 21
Q9YA39 PRELIMINARY; PRT; 105 AA.
ID Q9YA39;
AC Q9YA39;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE HYPOTHETICAL 11.3 KDA PROTEIN APE2100.
GN APE2100.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Akai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudooh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000063; BAA8111.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 105 AA; 11262 MW; B25D94F6CED05D87 CRC64;

Query Match 91.7%; Score 22; DB 17; Length 105;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 gtxxps 6
DB 15 GTSSPS 20

RESULT 22
O38165 PRELIMINARY; PRT; 111 AA.
ID O38165;
AC O38165;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024061; AAB83840.1; -.
DR HSSP; P05888; IAAF.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger.
FT NON_TER 1 111
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12601 MW; A3611A5B709F506 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
DB 91 GTTTS 96

RESULT 23
O38166 PRELIMINARY; PRT; 111 AA.
ID O38166;
AC O38166;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024062; AAB83841.1; -.
DR HSSP; P05888; IAAF.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.

```

```
KW Zinc-finger. 1 1
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12684 MW; FC3BEFE25A1B7EDB CRC64;

Query Match 91.7%; Score 22; DB 15; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 91 GTTTPS 96

RESULT 24
Q38169 PRELIMINARY; PRT; 111 AA.
AC Q38169;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RT "Drug resistance during indinavir therapy is caused by mutations in
RL the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024065; AAB83844.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger. 1 1
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12684 MW; 39C82CAA0BAFDDC CRC64;

Query Match 91.7%; Score 22; DB 15; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 91 GTTTPS 96

RESULT 25
Q38171 PRELIMINARY; PRT; 111 AA.
AC Q38171;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG PROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PATIENT 4;
RX MEDLINE=97404676; PubMed=9261388;
RA Zhang Y.M., Imamichi H., Imamichi T., Lane H.C., Falloon J.,
RT "Drug resistance during indinavir therapy is caused by mutations in
RL the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024065; AAB83844.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger. 1 1
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12684 MW; 39C82CAA0BAFDDC CRC64;

Query Match 91.7%; Score 22; DB 15; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 91 GTTTPS 96

RESULT 26
Q75698 PRELIMINARY; PRT; 112 AA.
AC Q75698;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG POLYPROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96213510; PubMed=8638406;
RA Barrie K.A., Perez E.E., Lamers S.L., Farmerie W.G., Dunn B.M.,
RA Sleasman J.W., Goodenow M.M.;
RT "Natural variation in HIV-1 protease, Gag p7 and p6, and protease
RT cleavage sites within gag/pol polyproteins: amino acid substitutions
RT in the absence of protease inhibitors in mothers and children infected
RL by human immunodeficiency virus type 1.";
RL Virology 219:407-416(1996).
DR EMBL; U53663; AAB05739.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Polyprotein; Zinc-finger.
FT NON_TER 112 112
SQ SEQUENCE 112 AA; 12895 MW; C97021F2A4407CA7 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 112;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 94 GTTTPS 99

RESULT 27
Q64053 PRELIMINARY; PRT; 135 AA.
AC Q64053;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
```

```
RA Vasudevachari M.B., Salzman N.P.;
RT "Drug resistance during indinavir therapy is caused by mutations in
RT the protease gene and in its Gag substrate cleavage sites.";
RL J. Virol. 71:6662-6670(1997).
DR EMBL; AF024067; AAB83846.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Zinc-finger. 1 1
FT NON_TER 111 111
SQ SEQUENCE 111 AA; 12482 MW; 653BE64427F9D8FC CRC64;

Query Match 91.7%; Score 22; DB 15; Length 111;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 91 GTATPS 96

RESULT 26
Q75698 PRELIMINARY; PRT; 112 AA.
AC Q75698;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE GAG POLYPROTEIN (FRAGMENT).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96213510; PubMed=8638406;
RA Barrie K.A., Perez E.E., Lamers S.L., Farmerie W.G., Dunn B.M.,
RA Sleasman J.W., Goodenow M.M.;
RT "Natural variation in HIV-1 protease, Gag p7 and p6, and protease
RT cleavage sites within gag/pol polyproteins: amino acid substitutions
RT in the absence of protease inhibitors in mothers and children infected
RL by human immunodeficiency virus type 1.";
RL Virology 219:407-416(1996).
DR EMBL; U53663; AAB05739.1; -.
DR HSSP; P05888; IAAF.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF00098; zf-CCHC; 2.
DR SMART; SM00343; Znf_C2HC; 2.
KW Polyprotein; Zinc-finger.
FT NON_TER 112 112
SQ SEQUENCE 112 AA; 12895 MW; C97021F2A4407CA7 CRC64;

Query Match 91.7%; Score 22; DB 15; Length 112;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6
   || ||
Db 94 GTTTPS 99

RESULT 27
Q64053 PRELIMINARY; PRT; 135 AA.
AC Q64053;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
```

DE ORF 5' OF FIBROBLAST GROWTH FACTOR RECEPTOR 1 FGFR-1.  
GN Mus sp.  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95100926; PubMed=7802632;  
RA Harada T., Saito H., Kouhara H., Kurebayashi S., Kasayama S.,  
RA Terakawa N., Kishimoto T., Sato B.;  
RT "Murine fibroblast growth factor receptor 1 gene generates multiple  
RT messenger RNAs containing two open reading frames via alternative  
RT splicing.";  
RL Biochem. Biophys. Res. Commun. 205:1057-1063(1994).  
DR EMBL: S74765; AAB32844.2; -  
DR MGD: MGI:95522; Fgfr1.  
KW Receptor.  
SQ SEQUENCE 135 AA; 14280 MW; 03906AD6DC40880F CRC64;  
  
Query Match 91.7%; Score 22; DB 11; Length 135;  
Best Local Similarity 66.7%; Pred. No. 1.8e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
Db 105 GTAAPS 110  
  
RESULT 28  
Q9D5C6 PRELIMINARY; PRT; 142 AA.  
ID Q9D5C6  
AC Q9D5C6  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE 4930465A12RIK  
GN 4930465A12RIK  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6J; TISSUE=TESTIS;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikola I., Pesole G., Quackenbush J.,  
RA Schirni L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamliya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
DR EMBL: AK015502; BAB29872.1; -  
DR MGD: MGI:1922239; 4930465A12RIK.  
SQ SEQUENCE 142 AA; 15397 MW; AF4B8868A17B745E CRC64;  
  
Query Match 91.7%; Score 22; DB 11; Length 142;

Best Local Similarity 66.7%; Pred. No. 1.9e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
Db 74 GTATPS 79  
  
RESULT 29  
Q9HBN7 PRELIMINARY; PRT; 156 AA.  
ID Q9HBN7  
AC Q9HBN7  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 16.2 KDA PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.O., Zhang P.P.,  
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,  
RA Yu J., Han L.H.;  
RT "Novel Human cDNA clones with function of inhibiting cancer cell  
RT growth.";  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
KW EMBL: AF218021; AAG17263.1; -  
SQ SEQUENCE 156 AA; 16178 MW; 183180BEC8F93E63 CRC64;  
  
Query Match 91.7%; Score 22; DB 4; Length 156;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6  
Db 10 GTAAPS 15  
  
RESULT 30  
Q9NWG3 PRELIMINARY; PRT; 157 AA.  
ID Q9NWG3  
AC Q9NWG3  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DE 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE HYPOTHETICAL 17.4 KDA PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE=EMBRYO;  
RA Isoqai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AK000905; BAA91418.1; -  
SQ SEQUENCE 157 AA; 17352 MW; 2B1C874775BC2D23 CRC64;  
  
Query Match 91.7%; Score 22; DB 4; Length 157;  
Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
Qy 1 gtxxps 6

Db 76 GTAAPS 81

```

RESULT 31
Q9VSO7 PRELIMINARY; PRT; 175 AA.
AC Q9VSO7;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE CG13309 PROTEIN.
GN CG13309.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flossler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003554; AAF50359.1; -.
DR FlyBase; FBgn0035933; CG13309.
SQ SEQUENCE 175 AA; 18038 MW; A094BA899ADFA1C7 CRC64;
```

Query Match 91.7%; Score 22; DB 5; Length 175;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
Db 98 GTSTPS 103

```

RESULT 32
O41995 PRELIMINARY; PRT; 177 AA.
AC O41995;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE COAT PROTEIN (FRAGMENT).
OS maize rayado fino virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Marafivirus.
OX NCBI_TaxID=59749;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UNITED STATES;
RX MEDLINE=98062136; PubMed=9400964;
RA Hammond R.W., Kogel R., Ramirez P.;
RT "Variability of geographically distinct isolates of maize rayado fino
RL virus in Latin America.";
RL J. Gen. Virol. 78:0-0(0).
DR EMBL; U97729; AAB96574.1; -.
DR InterPro; IPR000574; Tymo_coat.
DR Pfam; PF00983; Tymo_coat; 1.
FT NON_TER 1
SQ SEQUENCE 177 AA; 18918 MW; F1A95271D011795E CRC64;
```

Query Match 91.7%; Score 22; DB 12; Length 177;  
Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
Db 171 GTATPS 176

```

RESULT 33
Q9KY82 PRELIMINARY; PRT; 178 AA.
AC Q9KY82;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 19.7 KDA PROTEIN.
GN SKJ15.22.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashl H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL356813; CAB92613.1; -.
KW Hypothetical protein.
SQ SEQUENCE 178 AA; 19681 MW; 155BA91389A2DDE6 CRC64;
```

Query Match 91.7%; Score 22; DB 2; Length 178;  
 Best Local Similarity 66.7%; Pred. No. 2.3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 ||||  
 Db 2 GTTSPS 7

RESULT 34  
 Q9VSO6 PRELIMINARY; PRT; 182 AA.  
 ID Q9VSO6  
 AC Q9VSO6  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
 DE CGI3308 PROTEIN.  
 GN CGI3308.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.F., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 DR EMBL; AE003554; AAF50360.1; -  
 DR FlyBase; FBgn0035932; CGI3308.  
 SQ SEQUENCE 182 AA; 18808 MW; 64B9775C44A3AEE2 CRC64;

Query Match 91.7%; Score 22; DB 5; Length 182;  
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 ||||  
 Db 105 GTSPS 110

RESULT 35  
 Q9BZ94 PRELIMINARY; PRT; 186 AA.  
 ID Q9BZ94  
 AC Q9BZ94  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE ZINC FAMILY MEMBER 4 PROTEIN HZIC4.  
 GN ZIC4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MEDULLOBLASTOMA;  
 RA Warder D.E.;  
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF332509; AAK06845.1; -  
 DR HSSP; P08047; 1SP2.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam; PF00096; zf-C2H2; 1.  
 DR SMART; SM00355; Znf-C2H2; 3.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 2.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 SQ SEQUENCE 186 AA; 20189 MW; BF4C37753D290527 CRC64;

Query Match 91.7%; Score 22; DB 4; Length 186;  
 Best Local Similarity 66.7%; Pred. No. 2.4e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
 ||||  
 Db 67 GTATPS 72

RESULT 36  
 O82664 PRELIMINARY; PRT; 196 AA.  
 ID O82664  
 AC O82664;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE BLUE COPPER BINDING-LIKE PROTEIN (AT5G20230/F5024\_120) (PUTATIVE BLUE  
 DE COPPER BINDING PROTEIN).  
 GN AWI 32 OR ATBCD.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. LANSBERG ERECTA;  
 RA Honma T., Goto K.;  
 RT "Characterization of a wound-inducible Arabidopsis gene encoding a  
 RT protein homologous to blue copper binding proteins.";  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. LANSBERG ERECTA;  
 RA Honma T., Goto K.;  
 RT "The Arabidopsis floral homeotic gene PISTILLATA is regulated by  
 RT discrete cis-elements responsive to induction and maintenance  
 RT signals.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

RN SEQUENCE FROM N.A.  
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,  
 RA Ishida J., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
 RA Tshida L., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
 RA Ecker J.R.;  
 RT "Arabidopsis cDNA clones";  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN SEQUENCE FROM N.A.  
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,  
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,  
 RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,  
 RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E.,  
 RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,  
 RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,  
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;  
 RT "Full length cDNA of gene P5024.120/AT5g20230";  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y18227; CAA77089.1; -;  
 DR EMBL; AB035137; BAA86999.1; -;  
 DR EMBL; AY052681; AAK96585.1; -;  
 DR EMBL; AY034986; AAK59491.1; -;  
 DR HSSP; P29602; 1JER.  
 DR InterPro; IPR000923; Copper\_blue1.  
 DR InterPro; IPR003245; Cu\_bind\_like.  
 DR Pfam; PF02298; Cu\_bind\_like; 1.  
 DR ProDom; PD003122; Cu\_bind\_like; 1.  
 DR ProSITE; PS00196; COPPER\_BLUE; 1.  
 SQ SEQUENCE 196 AA; 20053 MW; 05100B50518F0A56 CRC64;

Query Match 91.7%; Score 22; DB 10; Length 196;  
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || || ||  
 Db 158 GTTTPS 163

RESULT 37  
 Q9F5I3 PRELIMINARY; PRT; 200 AA.  
 AC Q9F5I3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE R1ORF4 PROTEIN.  
 GN R1ORF4.  
 OS Agrobacterium rhizogenes.  
 OG Plasmid pRI1724.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=359;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N.,  
 RA Yoshida K.;  
 RT "The complete nucleotide sequence of a Ri (root inducing) plasmid  
 RT indicates its chimerical structure between ri and Sym plasmids.";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RA Moriguchi K., Maeda Y., Satou M., Kataoka M., Tanaka N., Yoshida K.;  
 RT "Analysis of unique variable region of a plant root inducing plasmid,  
 RT pRI1724, by the construction of its physical map and library.";  
 [2]

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;  
 RT "Genome structure of Ri plasmid (1): Construction of linking library  
 RT and physical map of pRI1724 in Japanese Agrobacterium.";  
 RL Nucleic Acids Symp. Ser. 39:189-190(1998).  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=MAFF03-01724;  
 RX MEDLINE=20241294; PubMed=10780382;  
 RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,  
 RA Yoshida K.;  
 RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and  
 RT its flanking regions of pRI1724 in Japanese Agrobacterium  
 RT rhizogenes.";  
 RL Nucleic Acids Symp. Ser. 42:67-68(1999).  
 CC -!- SIMILARITY: TO THE LYSR FAMILY OF TRANSCRIPTIONAL REGULATORS.  
 DR EMBL; AF002086; BAB16123.1; -;  
 DR InterPro; IPR000847; HTH\_LysR.  
 DR Pfam; PF00126; HTH\_1; 1.  
 KW DNA-binding; Plasmid; Transcription regulation.  
 SQ SEQUENCE 200 AA; 21360 MW; 6888FA7699BFCEAF CRC64;

Query Match 91.7%; Score 22; DB 2; Length 200;  
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || || ||  
 Db 177 GTAAPS 182

RESULT 38  
 Q9KQ11 PRELIMINARY; PRT; 202 AA.  
 AC Q9KQ11;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE OUTER MEMBRANE LIPOPROTEIN SLP, PUTATIVE.  
 GN VC1987.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae.";  
 RL Nature 406:477-483(2000).  
 DR EMBL; AE004273; AAF95135.1; -;  
 DR TIGR; VC1987; -;  
 KW Lipoprotein; Complete proteome.  
 SQ SEQUENCE 202 AA; 22661 MW; DA43B8A8D40DB691 CRC64;

Query Match 91.7%; Score 22; DB 16; Length 202;  
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || || ||

Db 134 GTTAPS 139

RESULT 39

Q9JLH5 ID Q9JLH5 PRELIMINARY; PRT; 218 AA.

AC Q9JLH5;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)

DE CDK5 ACTIVATOR-BINDING PROTEIN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=2018474; PubMed=10721722;

RA Ching Y.P., Qi Z., Wang J.H.;

RT "Cloning of three novel neuronal Cdk5 activator binding proteins.";

RL Gene 242:285-294(2000).

DR EMBL; AF177478; AAF60224.1; -

SQ SEQUENCE 218 AA; 24023 MW; 2C6750AABA0D0EDA CRC64;

Query Match 91.7%; Score 22; DB 11; Length 218;

Best Local Similarity 66.7%; Pred. No. 2.8e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 209 GTSSPS 214

RESULT 40

Q9RDK2 ID Q9RDK2 PRELIMINARY; PRT; 228 AA.

AC Q9RDK2;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)

DE PUTATIVE MEMBRANE PROTEIN.

GN SCC123.22C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Brown S.P., Harris D.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.;

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmid and a detailed genetic and physical map for

the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL136518; CAB66262.1; -

SQ SEQUENCE 228 AA; 24246 MW; 6217C85671050B30 CRC64;

Query Match 91.7%; Score 22; DB 2; Length 228;

Best Local Similarity 66.7%; Pred. No. 2.9e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 206 GTATPS 211

RESULT 41

Q9KXV0 ID Q9KXV0 PRELIMINARY; PRT; 233 AA.

AC Q9KXV0;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)

DE PUTATIVE SECRETED PROTEIN.

GN SCD95A.24.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Cerdeño A.M., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.;

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmid and a detailed genetic and physical map for

the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL357432; CAB93051.1; -

SQ SEQUENCE 233 AA; 23071 MW; D3E0BEE9A356083D CRC64;

Query Match 91.7%; Score 22; DB 2; Length 233;

Best Local Similarity 66.7%; Pred. No. 3e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6

|| ||

Db 52 GTATPS 57

RESULT 42

P79289 ID P79289 PRELIMINARY; PRT; 237 AA.

AC P79289;

DT 01-MAY-1997 (TReMBLrel. 03, Created)

DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE SPL TRANSCRIPTION FACTOR (FRAGMENT).

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI\_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=ENDOMETRIUM;

RX MEDLINE=20150569; PubMed=10687861;

RA Simmen R.C.M., Zhang X.L., Zhang D., Wang Y., Michel F.J.;

RA Simmen F.A.;

RT "Expression and regulatory function of the transcription factor Spl in

the uterine endometrium at early pregnancy: implications for

epithelial phenotype.";

RT Mol. Cell. Endocrinol. 159:159-170(2000).

DR EMBL: U57347; AAB39513.3; -.  
 DR HSP: P08047; ISP2.  
 DR InterPro: IPR000822; Znf-C2H2.  
 DR Pfam: PF00096; zf-C2H2; 3.  
 DR SMART: SM00355; Znf-C2H2; 3.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 3.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2\_2; 4.  
 KW DNA-binding; Metal-binding; Zinc-finger.  
 FT NON\_TER 1  
 FT NON\_TER 237  
 SQ SEQUENCE 237 AA; 25421 MW; C6950DB42912DAB6 CRC64;

Query Match 91.7%; Score 22; DB 6; Length 237;  
 Best Local Similarity 66.7%; Pred. No. 3e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 DB 204 GTATPS 209

RESULT 43  
 Q9L203 PRELIMINARY; PRT; 247 AA.  
 AC Q9L203;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE PUTATIVE BIOTIN SYNTHASE.  
 SC8E4.05C.

OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN=A3(2); PubMed=8843436;  
 RX MEDLINE=97000351; Denapaita D., Eichner A., Cullum J.,  
 RA Redenbach M., Kieser H.M., Hopwood D.A.;  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; ALI38661; CAB71805.1; -.  
 DR InterPro: IPR003784; BioY.  
 DR Pfam: PF02632; BioY; 1.  
 SQ SEQUENCE 247 AA; 24640 MW; 3A5EF99FDD7CF0CE CRC64;

Query Match 91.7%; Score 22; DB 2; Length 247;  
 Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 DB 150 GTAAPS 155

RESULT 44  
 Q9L126 PRELIMINARY; PRT; 253 AA.  
 ID Q9L126  
 AC Q9L126;

DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE PUTATIVE LIPOPROTEIN.  
 GN SC6D11.30.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL; ALI58061; CAB76353.1; -.  
 KW Lipoprotein.  
 SQ SEQUENCE 253 AA; 26600 MW; FECC3325A465D891 CRC64;

Query Match 91.7%; Score 22; DB 2; Length 253;  
 Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gtxxps 6  
 || ||  
 DB 122 GTASPS 127

RESULT 45  
 O43563 PRELIMINARY; PRT; 253 AA.  
 ID O43563;  
 AC O43563;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE ORCTL25 HYPOTHETICAL PROTEIN.  
 GN ORCTL25 OR BWR1B.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cooper P.R., Smilnich N.J., Day C.D., Nowak N.J., Reid L.H.,  
 RA Pearsall R.S., Reece M., Prawitt D., Landers J., Housman D.E.,  
 RA Winterpacht A., Zabel B.O., Pelletier J., Weissman B.E., Shows T.B.,  
 RA Higgins M.J.;  
 RL Genomics 0:0-0(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=98188297; PubMed=9520460;  
 RA Schiendacher C., Sabbioni S., Campi M., Veronese A., Bernardi G.,  
 RA Menegatti A., Hataida I., Mukai T., Ohashi H., Barbanti-Brodano G.,  
 RA Croce C.M., Negrini M.;  
 RT "Transcriptional map of 170-kb region at chromosome 11p15.5:  
 identification and mutational analysis of the BWR1A gene reveals the  
 presence of mutations in tumor samples.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:3873-3878(1998).  
 DR EMBL; AF037066; AAC04789.1; -.

DR EMBL; AF035407; AAC17497.1; -;  
SQ SEQUENCE 253 AA; 27218 MW; 183DE7B5C8A2255B CRC64;

Query Match 91.7%; Score 22; DB 4; Length 253;  
Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 240 GTSTPS 245

RESULT 46  
Q91170  
ID Q91170 PRELIMINARY; PRT; 254 AA.  
AC Q91170  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE PROBABLE THIOESTERASE.  
GN PA2411.  
OS Pseudomonas aeruginosa  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004668; AAG05799.1; -;  
DR InterPro; IPR001031; Thioesterase.  
DR Pfam; PF00975; Thioesterase; 1.  
KW Complete proteome.  
SQ SEQUENCE 254 AA; 27878 MW; 460C521723BD7403 CRC64;

Query Match 91.7%; Score 22; DB 16; Length 254;  
Best Local Similarity 66.7%; Pred. No. 3.2e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 108 GTAAPS 113

RESULT 47  
Q98TY7  
ID Q98TY7 PRELIMINARY; PRT; 275 AA.  
AC Q98TY7  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE MYOD.  
GN MYOD.  
OS Brachydanio rerio (Zebrafish) (Zebra danio).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
OC Cypriniformes; Cyprinidae; Danio.  
OX NCBI\_TaxID=7955;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gao J., Du S.;  
RT "Isolation and Characterization of Muscle-Specific Expression of

RT Zebrafish Myod Gene.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH AN OTHER  
CC BHLH PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
CC TRANSCRIPTION FACTORS.  
DR EMBL; AF318503; AAK06755.1; -;  
DR HSP; P10085; IMDY.  
DR InterPro; IPR002546; Basic.  
DR InterPro; IPR001092; HLH\_dim.  
DR InterPro; IPR003015; HLH\_Myc.  
DR Pfam; PF01586; Basic; 1.  
DR Pfam; PF00010; HLH; 1.  
DR SMART; SM00520; BASIC; 1.  
DR SMART; SM00353; HLH; 1.  
DR PROSITE; PS00038; HELIX\_LOOP\_HELIX; 1.  
KW DNA-binding  
SQ SEQUENCE 275 AA; 30918 MW; 0A395542F95B37D8 CRC64;

Query Match 91.7%; Score 22; DB 13; Length 275;  
Best Local Similarity 66.7%; Pred. No. 3.5e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 253 GTTAPS 258

RESULT 48  
Q9M0A9  
ID Q9M0A9 PRELIMINARY; PRT; 285 AA.  
AC Q9M0A9;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 30.9 KDA PROTEIN.  
GN AT4G30510.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lanar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,  
RA Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL161577; CAB79789.1; -;  
DR InterPro; IPR001680; WD40.  
DR SMART; SM00320; WD40; 1.  
KW Hypothetical protein; Repeat; WD repeat.  
SQ SEQUENCE 285 AA; 30935 MW; E8E7017F494969E1 CRC64;

Query Match 91.7%; Score 22; DB 10; Length 285;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 gtxxps 6  
|| ||  
Db 233 GTSSPS 238

RESULT 49  
Q9RNC9  
ID Q9RNC9 PRELIMINARY; PRT; 288 AA.  
AC Q9RNC9;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
GN VIRB9 HOMOLOG.  
OS Bartonella henselae (Rochalimaea henselae).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Bartonellaceae; Bartonella.  
OX NCBI\_TaxID=38323;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HOUSTON-1;  
RX MEDLINE=20338084; PubMed=10882236;  
RA Padmalayam I., Karem K., Baumstark B., Massung R.;  
RT "The gene encoding the 17-kDa antigen of Bartonella henselae is  
RT located within a cluster of genes homologous to the virB virulence  
RT operon.";  
RL DNA Cell Biol. 19:377-382(2000).  
DR EMBL: AF182718; AAF0947.1; -. B94F15C58F38D489 CRC64;  
SQ SEQUENCE 288 AA; 32120 MW; B94F15C58F38D489 CRC64;  
  
Query Match 91.7%; Score 22; DB 2; Length 288;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 gtxxps 6  
|| ||  
Db 271 GTTSPS 276  
  
RESULT 50  
Q9DWD6  
ID Q9DWD6 PRELIMINARY; PRT; 288 AA.  
AC Q9DWD6;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)  
DE PR58.  
GN R58.  
OS Rat cytomegalovirus (strain Maastricht).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Muromegalovirus.  
OX NCBI\_TaxID=79700;  
RN [1]  
RP SEQUENCE OF 1-168 FROM N.A.  
RC STRAIN=MAASTRICHT;  
RX MEDLINE=96335691; PubMed=8757999;  
RA Beuken E., Slobbe R., Bruggeman C.A., Vink C.;  
RT "Cloning and sequence analysis of the genes encoding DNA polymerase,  
RT glycoprotein B, ICP18.5 and major DNA-binding protein of rat  
RT cytomegalovirus.";  
RL J. Gen. Virol. 77:1559-1562(1996).  
RN [2]  
RP SEQUENCE OF 167-288 FROM N.A.  
RC STRAIN=MAASTRICHT;  
RX MEDLINE=98033207; PubMed=9367384;  
RA Vink C., Beuken E., Bruggeman C.A.;  
RT "Cloning and functional characterization of the origin of lytic-phase  
RT DNA replication of rat cytomegalovirus.";  
RL J. Gen. Virol. 78:2963-2973(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAASTRICHT;  
RX MEDLINE=20366325; PubMed=10906222;  
RA Vink C., Beuken E., Bruggeman C.A.;  
RT "Complete DNA sequence of the rat cytomegalovirus genome.";  
RL J. Virol. 74:7656-7665(2000).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MAASTRICHT;  
RX MEDLINE=20473137; PubMed=11018281;  
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;  
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a

RT spliced transcript.";  
RL Virus Res. 69:119-130(2000).  
DR EMBL: AF232689; AAF99154.1; -.  
SQ SEQUENCE 288 AA; 30131 MW; C8F592D7837C9A34 CRC64;  
  
Query Match 91.7%; Score 22; DB 12; Length 288;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 gtxxps 6  
|| ||  
Db 247 GTATPS 252  
  
Search completed: September 3, 2002, 09:29:29  
Job time: 236 sec

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**